EVALUATION OF THE UNIVERSITY OF CANTERBURY AUDITORY-VISUAL MATRIX SENTENCE TEST

A thesis submitted in partial fulfilment of the requirements for the Degree of

Master of Audiology

in the Department of Communication Disorders

at the University of Canterbury

J. M. Stone

University of Canterbury

2016

ACKNOWLEDGEMENTS

Completing this thesis would not have been possible without the help of a few key people.

Firstly I would like to thank my primary supervisor, and the brain behind the intricacies of this project, Dr. Greg O'Beirne. I am so grateful for all of the time you gave to this project and for the endless knowledge and encouragement you provided throughout the journey that is writing a thesis. Your enthusiasm for all things 'Audiology' is an inspiration.

To my co-supervisor, Dr. Rebecca Kelly-Campbell, I cannot thank you enough for the time you spent assisting me with this project. I am so appreciative of your support and for all of the feedback you have given me, despite your already overwhelming workload. The dedication you show your students is truly amazing.

I would also like to extend my sincerest thanks to all of the MAud academic and clinical staff for your unwavering support throughout the programme.

I would like to express genuine thanks to each of the participants that willingly gave their time to this project – it simply would not have been possible without you.

To my classmates, congratulations! I could not of hoped to go through my training with a better group of people, thank you for making this process so enjoyable. I wish each of you all the best for the bright futures that I know lay ahead.

To my friends, thank you for always caring enough to ask the dreaded "how is study going" question and listening to my groans and grumbles as I attempted to answer throughout the challenges that these last six years of study have presented me with. You girls mean the world.

To my partner, Jack, there are no words to express how grateful I am for everything you do. Thank you for your love and patience on those early mornings that you know I find so difficult and for grounding me in the moments of sheer panic over the last two years. I am so thankful to have had you by my side throughout this journey.

Lastly, I would like to acknowledge my family. Dad, I am so thankful to you for always encouraging us to succeed. Your support over the last few years has helped more than you could know. To my sister, Lois, I really appreciate all of the guidance you have given me over the years and for showing me the ropes where this 'studying business' was concerned. I cannot thank you enough for always being there to listen and understand when things were difficult and to celebrate with me when the hard work paid off. I couldn't of done it without you. And last, but by no means least, Mum, I'd like to thank you for your unconditional love through it all. Your endless encouragement and support while I pursued my direction could not go unrecognised. I dedicate this, my final piece, to you.

ABSTRACT

Speech recognition tests are an important tool in audiology, providing information regarding an individual's communication deficits beyond that revealed by the audiogram. With the aim of providing an accurate representation of such deficits faced in the real world, the University of Canterbury Auditory-Visual Matrix Sentence Test (UCAMST) was developed in New Zealand (NZ) English (O'Beirne, Trounson, McClelland, Jamaluddin, & Maclagan, 2015; Trounson, 2012). While international versions of this measure exist in various languages and dialects of English, in order to preserve the validity of the measure, development of a NZ English version was warranted. The current study sought to evaluate the lists generated for use in both the auditory and auditory-visual modalities to establish the equivalence of the lists and conditions of the UCAMST. Further, in order to determine whether the UCAMST sentence stimuli were in accord with international standards, evaluation across previous versions was conducted. Evaluation of the UCAMST stimuli with 42 participants with normal hearing (NH) revealed that while some of the lists were equivalent to one another, the conditions were not. Further, results showed the UCAMST to differ from international versions. These findings, while encouraging in part, require the attention of future research as equivalence is of critical importance in the ability to compare results across sessions and clinics. Overall, this research constituted one study in a series of many aimed at progressing the UCAMST towards implementation in the audiological test battery in NZ.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABG	Air-Bone Gap
AM	Amplitude Modulation
ANOVA	Analysis of Variance
BM	Basilar Membrane
dB	Decibels
dB HL	Decibels Hearing Level
dB SNR	Decibels Signal-to-Noise Ratio
dB SPL	Decibels Sound Pressure Level
CVC	Consonant-Vowel-Consonant
DANTALE II	Danish Matrix Sentence Test
EU	European Union
HA	Hearing Aid
HAPI	Hearing Aid Performance Inventory
HI	Hearing Impairment
HINT	Hearing in Noise Test
Hz	Hertz
L_{mid}	50% correct point shown on an intelligibility function
MST	Matrix Sentence Test
MTA	Motor Trade Association
NH	Normal Hearing
NZ	New Zealand
NZDTT	New Zealand Digit Triplet Test

NZ HINT	New Zealand Hearing in Noise Test
PI	Performance-Intensity
PPMST	Polish Pediatric Matrix Sentence Test
QuickSIN	Quick Speech in Noise
RM-ANOVA	Repeated-Measures Analysis of Variance
$s50_{test}$	Test-Specific Slope
$s_{sentence}$	Sentence-Specific Slope
s_{word}	Word-Specific Slope
SD	Standard Deviation
SNHI	Sensorineural Hearing Impairment
SNR	Signal-to-Noise Ratio
SPIN	Speech Perception in Noise
SPSS	Statistical Package for the Social Sciences
SRT	Speech Recognition Threshold
sSRT	Sentence Speech Recognition Threshold
TM	Tympanic Membrane
UCAMST	University of Canterbury Auditory-Visual Matrix Sentence Test
UCAST	University of Canterbury Adaptive Speech Test
WHO ICF	World Health Organisation International Classification of Functioning, Disability and Health

DEFINITIONS

The nomenclature used throughout this thesis attempted to remain consistent with that of the model provided by the World Health Organisation's International Classification of Functioning, Disability and Health (WHO ICF; World Health Organisation, 2001).

In accordance with this aim, the term "hearing loss" was replaced with the term "hearing impairment" in order to acknowledge the multifaceted nature of hearing impairment.

Similarly, in order to conform to the WHO ICF principle of universality, when referring to individuals with a hearing impairment the phrase "hearing impaired persons" was not used so as to avoid the labelling of individuals with this disability as a separate social group.

Additionally, the term "client" replaced the term "patient" in an attempt to portray the client-centred approach of the WHO ICF model. The philosophy underlying this approach is that the partnership between the client and the clinician promotes client participation in the selection of treatment and rehabilitation regimes, thus improving outcomes.

CHAPTER ONE

1.1 Background

Hearing impairment (HI), an invisible disability that can lead to disruption in communication and wellbeing, is the most prevalent sensory disability affecting human populations (Bird & O'Beirne, 2015; Mathers, Smith & Concha, 2000; Olusanya, Neumann & Saunders, 2014). Specifically, it has been estimated that HI affects 25-80% of adults over the age of 65 years (Newman & Sandridge, 2004). The negative effects of this disability can be profound and extend beyond that of auditory impairment, including communication difficulties, social and emotional isolation, physical health concerns and negative perceptions of overall quality of life (Kelly-Campbell & Lessoway, 2015; Mulrow et al*.,* 1990; Newman & Sandridge, 2004). Such negative outcomes reach beyond the individual and can have a significant impact on relationships among family members as well as with significant others (Kelly-Campbell & Lessoway, 2015). In order to reduce such consequences, the importance of gaining audiologic information that describes the individual's experience cannot be understated with regards to the rehabilitation of those with HI.

Currently, the most common rehabilitative approach used to address HI in adults is through the prescription of hearing aids (HA) (Chisolm et al., 2007; Kelly-Campbell & Lessoway, 2015). However, despite the well-documented evidence surrounding the perceived benefit of HA use, determining whether an individual is a suitable candidate poses a number of complexities (Cook & Hawkins, 2006). Alongside factors such as the motivation to use HAs and concerns regarding cosmetics and cost, the individual's perceived deficit is of pivotal concern (Mulrow et al., 1990). As with all health concerns, each individual's experience is unique and HI is no exception to this. The psychosocial effects of a HI for an individual may be considered to be severe, despite the degree of HI measured on the audiogram being relatively mild (Mulrow et al., 1990). Thus, gaining information relating to the effects of HI in the real world as well as the benefit that may be expected from HA use is critical to the work of rehabilitation audiologists.

In order to gain a more comprehensive understanding of such deficits, speech recognition tests are generally performed during an audiologic assessment. The results of these measures provide information regarding the individual's ability to detect and understand speech stimuli (Mendel, 2008). These tests therefore serve as important tools in gaining an understanding of the communication difficulties faced in various acoustic environments, and in providing direction regarding hearing rehabilitation (Dietz et al., 2014; Ozimek, Warzybok & Kutzner, 2010). A wide range of speech recognition measures exist, and continue to be developed, today, and are the foundation of this thesis. Following the development a new speech recognition measure in NZ, this project aims to evaluate this new tool in the hope of progressing it towards routine use in research and audiologic assessment in NZ.

1.2 Hearing Impairment

1.2.1 Anatomy of Hearing

In order to discuss hearing assessment, the auditory system and how it normally functions should first be described. The human auditory system can be divided into four main parts – the outer ear, the middle ear, the inner ear and the auditory neural pathway – that function in synchrony to enable hearing (Gates & Mills, 2005). The outer ear consists of the pinna, the most visible portion of the ear, and the external auditory meatus (i.e. the ear canal) which together act as a resonator to enhance the transmission of sound to the corresponding sections of the ear (Bess & Humes, 2008; Gates & Mills, 2005). The tympanic membrane (TM) separates the outer ear and middle ear, which is comprised of the tympanic cavity and the ossicular chain (Bess & Humes, 2008). The ossicular chain is formed by three bones (i.e. the ossicles) – the malleus, incus and stapes – which function to transfer air vibrations into the fluid-filled inner ear where they can be converted into chemical and electrical energy (Gates & Mills, 2005; Hall, 2014). The inner ear begins at the oval window, to which the broad base of the stapes (i.e. the footplate) is attached, and includes the sensory organ of hearing (i.e. the cochlea) and the organs of balance – the semicircular canals, the utricle, and the saccule (Bess & Humes, 2008; Gelfand, 2010). Despite the importance of the balance (or vestibular) system, the current research is focussed toward the hearing mechanism and thus the vestibular system will not be referred to again herein. Within the cochlea are three fluid-filled compartments: the perilymph-filled scala vestibuli and scala tympani, and the endolymph-filled scala media. Scala media lies between scala vestibuli and scala tympani, separated by Reissner's membrane and the basilar membrane (BM) respectively (Gelfand, 2010). Scala media contains the organ of Corti, where the sensory receptors for hearing (i.e. the hair cells) are located (Gelfand, 2010). The human cochlea contains 12,000 outer hair cells, which are situated across three rows at the basal turn, becoming four or five at the second apical turn, and 3,500 inner hair cells that lie in a single row (Donkelaar & Kaga, 2011). As sound waves enter the ear, the TM is set into vibrating movements that are sent to the inner ear via the corresponding motion of the ossicular chain (Donkelaar & Kaga, 2011). Once in the cochlea, sound vibrations produce small waves in the inner ear fluids causing displacement of the BM (Donkelaar & Kaga, 2011; Hall, 2014). The motion of the BM puts force on the stereocillia attached to the tip of each hair cell, allowing the influx of positive ions, which depolarises them and causes the inner hair cells to release a neurotransmitter, which in turn stimulates the auditory nerve fibres (Hall, 2014). From here, the signal is sent along the auditory neural pathway of the brain where it can be interpreted.

1.2.2 **Anatomy of Hearing Impairment**

If abnormalities occur within any of the aforementioned structures and/or processes, HI is the likely result. There are two types of HI that can be distinguished based on the location at which the problem occurs – conductive and sensorineural (Zeng & Liu, 2006). Conductive HI occurs when a complication arises in the outer or middle ear that physically interrupts the passage of sound to the cochlea (Donkelaar & Kaga, 2011). A variety of disorders can lead to a conductive HI, however, most are treatable through medical or surgical intervention and thus this form of HI is considered to be temporary in nature (Bess & Humes, 2008). In contrast, sensorineural HI (SNHI) is a consequence of pathology in the cochlea or the central connections to the cochlea nerve (Bess & Humes, 2008; Donkelaar & Kaga, 2011). SNHI is common and can arise as a result of a wide variety of conditions including tumours, infection, ageing and exposure to excessive noise or ototoxic medication (i.e. medication with known harmful side effects on the auditory system) (Donkelaar & Kaga, 2011). The most common cause of SNHI is the loss of sensory hair cells and, given the inability for hair cells to regenerate, the effects of this type of HI are permanent (Gates & Mills, 2005; Welberg, 2008). The effects of SNHI are extensive including the attenuation and distortion (i.e. the loss of clarity) of some, or all, sounds in addition to the numerous psychosocial effects previously mentioned (Kelly-Campbell & Lessoway, 2015; Mulrow et al., 1990; Newman & Sandridge, 2004; Plomp, 1978). Thus, based on its life-long nature and the large proportion of the population that are affected by this disability, SNHI is a major health concern (Schmiedt, 2010).

1.2.3 **Detection of Hearing Impairment**

Gaining information regarding a listener's hearing sensitivity during an audiologic assessment is typically achieved through undertaking pure tone audiometry. This procedure entails the client listening for pure tones of varying intensity and frequency in order to determine the lowest level at which the listener will detect a stimulus 50% of the time (i.e. the threshold [in dB HL]) (Valente, 2009). Conventional pure tone audiometry generally assesses a listener's threshold at octave frequencies between 250 Hz – 8000 Hz (Schlauch & Nelson, 2009). The results obtained through this procedure are plotted onto an audiogram to enable quantification of the type, severity and configuration of the HI (Schlauch & Nelson, 2009). As noted, the type of HI is inferred based on the site of lesion. The configuration of the HI refers to the shape of the HI depicted on the audiogram, which can be defined as: "flat", "gradually falling", "precipitously falling", "rising", "peaked", "trough", or "notched" (Lloyd & Kaplan, cited in, Schlauch & Nelson, 2009, p. 41). Last, the severity of the HI classifies the degree to which hearing sensitivity has reduced. Clark's (1981) classification system is utilised in NZ and categorises HI as being slight (16-25 dB HL), mild (26-40 dB HL), moderate (41-55 dB HL), moderately-severe (56-70 dB HL), severe (71-90 dB HL) or profound (≥ 91 dB HL). The degree of HI is determined by averaging the thresholds across the following frequencies: 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz.

While pure tone audiometry is imperative to the audiologic diagnostic assessment, the information obtained is most valuable when implemented alongside complementary measures, such as speech audiometry. As described, speech audiometry yields information beyond that of the audiogram, thus furthering the inferences and recommendations that can be made from the assessment results.

1.3 Speech Audiometry

Speech audiometry is an integral component of the audiologic test battery with its contribution to describing auditory function recognised for more than 50 years (Hall, 2008; Mendel, 2008; Talbott & Larson, 2008). As described, the primary aim of conducting speech audiometry is to obtain information regarding a listener's sensitivity to, and understanding of, speech sounds (i.e. speech recognition) through using speech as the target stimuli (Mendel, 2008). Thus when used alongside pure tone audiometry, the audiologist has the capacity to gain a more complete index of "hearing" and the level of dysfunction faced by the individual in daily life (Hall, 2008; Hamid & Brookler, 2006; Mendel, 2008). Due to this, the clinical applications of speech audiometry are vast, ranging from the diagnosis of auditory processing deficits to HA candidacy, hence the value placed on its use (Hall, 2008).

Despite the fundamental role of speech audiometry in clinical audiology, there are number of aspects of its use that require careful consideration in order to maximise the information that can be obtained. Such aspects include the presence or absence of competing background noise (i.e. masking noise), the method through which the masking noise is applied, and the chosen target stimuli. These considerations will be critically appraised in the following sections, based on the abundance of literature available in this area.

1.3.1 Speech Audiometry in NZ: Speech Recognition in Quiet

In NZ, speech recognition is commonly assessed through the use of monosyllabic word lists presented in quiet, such as the Meaningful Consonant-Vowel-Consonant (CVC; Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy, Arlington,

& Johnstone, 2000) words. The items are presented in lists, often after the carrier phrase “say [the word] _____”, whereby a listener is required to repeat the identified word following each presentation. Performance is measured by calculating the number of phonemes correctly identified in each word, enabling the percentage of correct responses across the list to be calculated. Three word lists are typically completed for each ear, at differing intensity levels (in dB HL) in order to describe a performance-intensity (PI) function (McArdle & Chisolm, 2009). A PI function depicts the relationship between the speech recognition probability and the average speech amplitude, and is utilised as a method of speech recognition threshold (SRT) – the level at which a listener correctly identifies 50% of the stimuli presented – estimation (Boothroyd, 2008). The other aspect of a listener’s performance that is described by a PI function is the PB_{max}, the presentation level at which the listener is considered to achieve maximum performance (Boothroyd, 2008).

The information derived from a PI function has a number of applications in clinical audiology. First, the function generated from a given listener can be compared to a normative curve whereby performance can be assessed in relation to the performance expected from listeners with NH. Second, the estimates of SRT obtained can be used as a method in crosschecking the validity of pure tone thresholds (Mendel, 2008). Last, and of particular importance to the clinical utility of the PI function, is the ability to employ phoneme scoring (Boothroyd, 2008). Phoneme scoring measures a listener’s performance not as the percentage of words correctly identified, but as a percentage of the constituent vowels and consonants recognised (Boothroyd, 2008). This method has several advantages, as compared to word scoring, including the ability to test an increased number of test items in a relatively short timeframe, which subsequently yields an increase in the measure’s test-retest reliability (Gelfand, 1998).

In addition to this, this method ensures that a listener's overall performance is less influenced by their vocabulary knowledge, which consequently is thought to provide a more valid measure of auditory resolution (Olsen, Van Tasell & Speaks, 1997).

Thus, based on the extent of the information derived from measures such as the CVC word lists, and the efficiency of such tools, the rationale behind the extensive use of word recognition tests in NZ clinics becomes apparent. Contrary to current practices however, empirical evidence has identified a number features that may advise reconsideration of the sole use of such measures in the audiologic test battery.

1.3.2 **Disadvantages of Measures of Speech Recognition in Quiet**

Despite the notable uses of word recognition tests presented in quiet, the shortcomings of such measures are also widely acknowledged (Orchik, Krygier & Cutts, 1979; Wilson, McArdle & Smith 2007a). In clinical practice the conditions under which these tests are performed are unlike those encountered in the real world. The speech stimuli for these measures are presented in isolation, with no contextual cues, in the absence of any competing background noise. The premise behind this format is that it may capture the problem of audibility more accurately, as compared to other test formats that may be confounded by factors such as a listener's ability to make use of contextual cues (Wilson et al., 2007a). In contrast to this premise however, the most frequent complaint brought to audiologists surrounds the communication difficulties faced when in the presence of competing background noise (Beattie, Barr & Roup, 1997; Dirks, Morgan & Dubno, 1982). Further, despite the recurrence of this issue, the typical test battery employed in clinics across NZ, and internationally, does not implement measures that directly assess such concerns. Due to the inability to predict speech recognition performance in noise from assessments taken in ideal acoustic conditions, providing clients and their family members with a realistic index of the problem is therefore hindered by the use of such measures (Beattie et al., 1997).

A major consequence of the inability to capture a listener's communication difficulty from the speech recognition tests currently employed is the inability to outline the benefits that may be expected following the dispense of amplification (Beattie et al., 1997). While it may seem reasonable to expect improved speech recognition to be a given outcome following the use of HAs, research has demonstrated that the extent of the communication handicap faced in daily life cannot be determined solely from measures of the loss of hearing sensitivity and disruption to speech understanding in quiet (Carhart & Young, 1976). Moreover, it has been suggested that HAs may exacerbate the problems associated with background noise for some individuals (Carhart & Young, 1976). Therefore, the need to establish the difficulty faced by a client in conditions that typify the complex listening environments of everyday life cannot be understated.

In addition to this issue, it is also important that diagnostic tests are sufficiently sensitive to discriminate between listeners with varying degrees of HI and those with NH. Research in this area has provided evidence that the deficits experienced by individuals with a mild HI may not be accurately reflected by performance on monosyllabic word measures in quiet (Beattie et al., 1997). It is thought that such tasks may be too straightforward to separate those with NH and the difficulties faced by those with a mild HI (Beattie et al., 1997).

Thus the almost exclusive use of word recognition measures presented in quiet in clinical audiology may significantly limit the inferences regarding a client's difficulty in real world situations and the rehabilitation suggestions that can be made by audiologists. Accordingly, while efficiency is vital in clinical settings, where time constraints exist, many researchers have proposed that speech recognition measures in noise, in addition to those employing sentence stimuli, may be more valuable to clinical practice (Beattie et al., 1997; Carhart & Young, 1976; Dirks et al., 1982).

1.4 Measures of Speech Recognition Presented in Noise

In order to address the believed disadvantages of speech recognition measures designed for use in quiet, measures of speech recognition in noise were developed (Taylor, 2003). It has been suggested that the use of both word and sentence stimuli presented in noise provide more powerful information regarding the deficits faced by the client in real world listening environments (Grunditz & Magnusson, 2013). Such information is thought to aid the clinician's ability to predict candidacy for various amplification methods and in counselling clients and their family members regarding the benefits and drawbacks of such methods in order to establish realistic expectations (Humes, 1999; Taylor, 2003). There are a wide range of speech in noise measures available for clinical use that differ with regards to a number of factors, including procedural parameters, such as the type of interfering masking noise or stimulus used, and presentation modes, such as the treatment of the stimulus or noise (Arlinger, 1998; Taylor, 2003; Wagener & Brand, 2005).

1.4.1 **Psychophysical Parameters**

As with measures of speech in quiet, performance on a speech in noise task is typically indicated by a listener's SRT (Brand & Kollmeier, 2002). In noise, however, the SRT is derived from a psychometric function that represents the relationship between a listener's performance (i.e. the percent correct score) on a psychophysical task and some physical aspects of the stimuli (i.e. the signal-to-noise ratio [SNR]) (MacPherson & Akeroyd, 2014). Psychometric functions are typically sigmoid-shaped (i.e. 's'-shaped) and are often summarised by two key parameters: the threshold – the stimulus level required to obtain 50% correct – and the slope – the rate at which performance increases with changes in the stimulus (Gilchrist, Jerwood & Ismaiel, 2005; MacPherson & Akeroyd, 2014). Figure 1 depicts the typical form of a psychometric function.

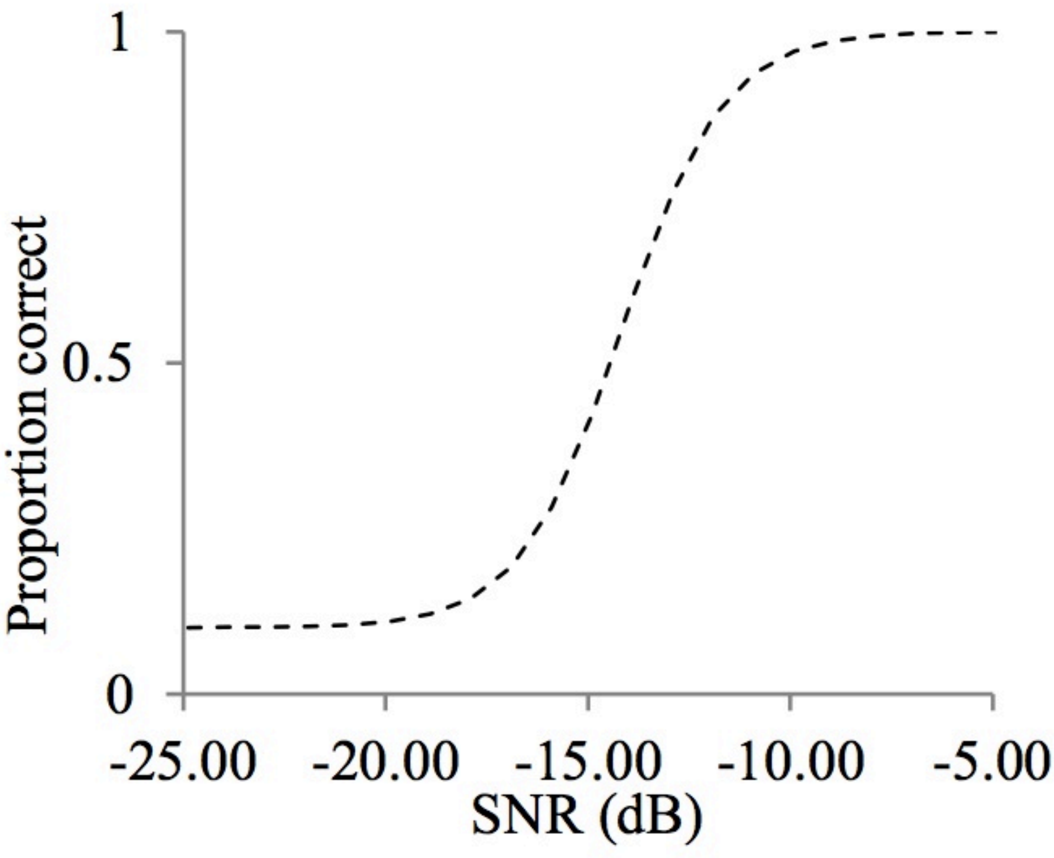

Figure 1. The typical shape associated with the psychometric function measuring the proportion of correct responses (%) against the SNR (dB). *Image retrieved from McClelland (2014, p. 12).*

With regards to speech in noise tasks, the slope is of critical importance as it determines the increase in perceptual benefit that a listener is likely to gain from small changes in the SNR (MacPherson & Akeroyd, 2014). Therefore, a steep psychometric function indicates that a small change in SNR leads to a large increase in intelligibility (MacPherson & Akeroyd, 2014). Conversely, the opposite is true for a shallow slope in that the same SNR improvement would lead to a smaller change in perceptual benefit. This conception is exemplified in Figure 2. The information obtained from the slope of a psychometric function has been postulated to be beneficial to the work of rehabilitation audiologists. It is thought that quantifying the amount of perceptual benefit a listener is likely to gain from the changes in SNR provided by a HA may assist in determining the recommendations to be conveyed to a given client (MacPherson & Akeroyd, 2014).

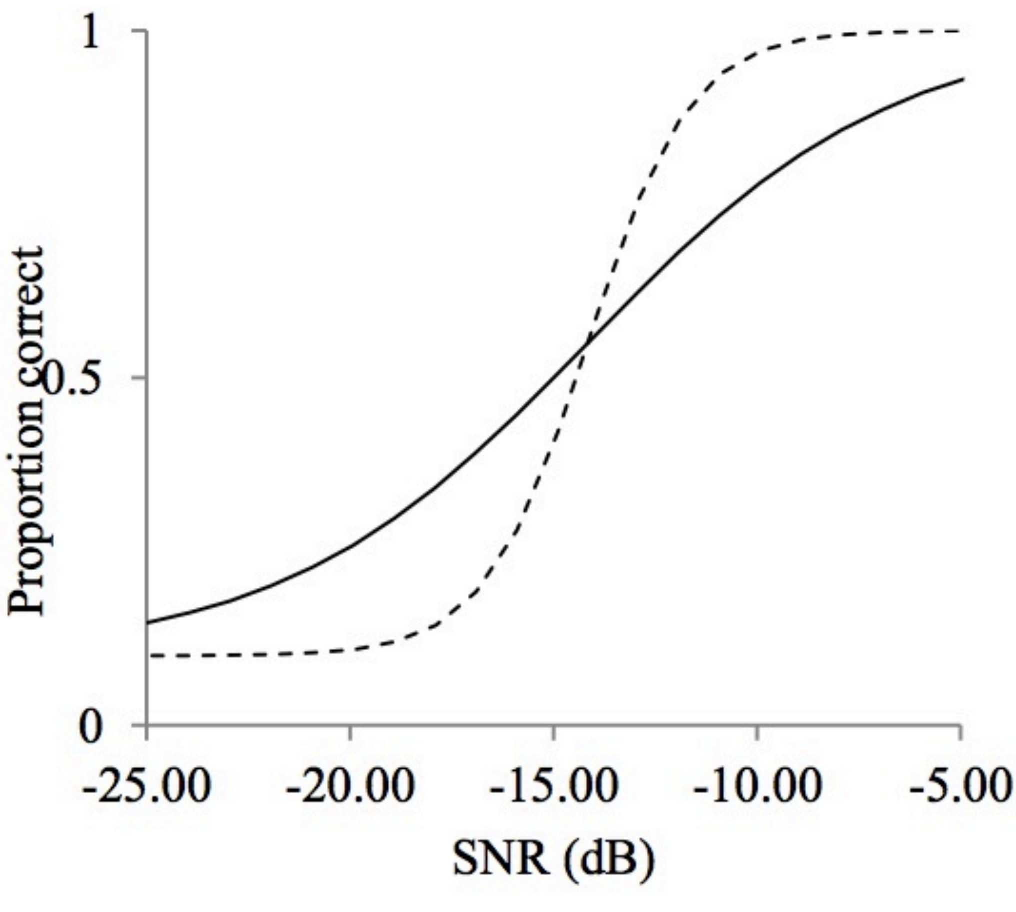

Figure 2. Comparison of psychometric functions with a steep (dashed line) and shallow (solid line) slope. *Image retrieved from McClelland (2014, p. 13)*

Additionally, a steep slope is thought to signify the sensitivity of a measure (Ozimek et al., 2010). A highly sensitive measure is considered to be desirable in that a listener's SRT can accurately be determined in a relatively small number of trials (Francart, van Wieringen & Wouters, 2011).

In summation, the psychometric function is considered to be a useful tool in speech audiometry as critical information with regards to the benefit that may be offered by various HA technologies can be obtained (MacPherson & Akeroyd, 2014). Such information may not only be beneficial in predicting a client's HA outcomes, but may also assist in counselling the client regarding their expectations of a HA and the benefit that they are likely to perceive. Further, based on the ability to derive information regarding the properties of a speech measure, inspection of a test's psychometric function may be valuable in the selection of a tool that complements the goals of the assessment (MacPherson & Akeroyd, 2014).

1.4.2 Selection of Masking Noise

In order to simulate a realistic listening environment, speech recognition measures can be administered in the presence of an acoustic masker (Francart et al., 2011). However, contingent on the objective of the test and the information sought, a certain type of masker may be more appropriate than another, and thus is an aspect of speech audiometry that requires careful consideration (Francart et al., 2011). The types of making noise typically utilised for speech recognition measures are continuous speech-shaped noise and multi-talker babble noise (Killion et al., 2004). The advantages of each method have been well documented in the literature. First, it has been suggested that speech-shaped noise has reduced variability, as compared to babble noise, thus allowing control to be easily achieved and thus the reproducibility of results to be improved as a consequence (Bacon, Opie & Montoya, 1998; Killion, Niquette, Gudmundsen, Revit & Banerjee, 2004). Based on such merits, continuous speech-shaped noise is likely to be a valuable tool when used in research settings.

Conversely, babble noise is thought to represent everyday speech-in-noise more accurately, and thus have higher face validity (Killion et al., 2004). Further, due to the fluctuating nature of babble noise it is thought to exhibit greater amplitude modulation (AM) than steady background noise (Bacon et al., 1998; Hopkins & Moore, 2009). AM is the gradual change in amplitude that provides NH subjects with a "glimpse" of the target signal, a phenomenon termed "masking release" (Hopkins & Moore, 2009; Howard-Jones & Rosen, 1993). For listeners with HI however, masking release is generally small or absent, and therefore it is possible that use of this masking noise may more closely reflect the difficulties with listening in background noise commonly reported by clients (Bacon et al., 1998; Hopkins & Moore, 2009). Therefore, the use of babble noise may be best suited to clinical assessments (Francart et al., 2011). Based on the literature, the importance of considering the merits of the masking noise when implementing a speech measure becomes apparent.

1.4.3 Fixed and Adaptive SNR Measures

Speech recognition tasks that determine a percent correct score at a fixed SNR are referred to as fixed SNR measures. The development of these measures had the primary aim of enabling the listening conditions typically faced in the real world to be approximated more closely (Taylor, 2003). Based on the premise that face-to-face communication becomes more difficult as the background noise increases, Pearsons, Bennett and Fidell (cited in Taylor, 2003) postulated that when the background noise was 55 dB SPL, the average intensity of the speech was 61 dB SPL (i.e. +6 dB SNR). Further, as the background noise increased to 65 dB SPL, the average speech was determined to be 68 dB SPL (i.e. +3 dB SNR), and likewise when the background noise was 75 dB SPL, the average speech was 74 dB SPL (i.e. -1 dB SNR) (Pearsons et al., cited in Taylor, 2003). Thus, it is believed that use of a fixed SNR test has the ability to simulate the following listening conditions – ‘relatively easy’, ‘moderately difficult’, and ‘challenging’, respectively (Taylor, 2003).

An example of a commercially available fixed SNR measure is the Speech Perception in Noise (SPIN; Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984) test. This measure utilises multi-talker babble noise and requires the listener to recall the final word, a monosyllabic noun, of the sentence stimuli presented (Bilger et al., 1984). Sentences are divided into equal groups of high-predictability and low-predictability and scoring is completed as a correct word percentage, which can be evaluated with regards to the predictability of the sentence (Bilger et al., 1984).

Fixed SNR measures, such as the SPIN (Bilger et al., 1984), are thought to be advantageous to clinical practice due to the ability to conduct testing in both the aided and unaided conditions. Testing in the aided and unaided conditions is thought to be beneficial in that use of the results from before and after a HA fitting is believed to provide evidence of HA benefit (Taylor, 2003). Such applications of these measures should be interpreted with caution however, based on the limited evidence supporting such claims. Much research has been conducted with the aim of revealing various speech measures to be sensitive enough to demonstrate objective HA benefit (Mendel, 2007; Parving, 1991). While such investigations have supported the use of aided and unaided testing in the HA evaluation process, the need for further research is warranted before concluding their ability to effectively capture an individual’s handicap (Niemeyer, 1976). For example, Mendel (2007) examined the use of the SPIN (Bilger et al., 1987) in addition to adaptive speech measures, detailed below, including the Hearing in Noise Test (HINT; Nilsson, Soli & Sullivan, 1994) and the

Quick Speech In Noise (QuickSIN; Killion et al., 2004) test, as an objective measure of HA benefit alongside subjective measures. The subjective measure chosen for this investigation was the Hearing Aid Performance Inventory (HAPI; Walden, Demorest & Hepler, 1984) due to the ability to approximate speech perception in various situations (Mendel, 2007). Significant results were revealed between the HAPI and all objective measures, except the noise condition for the HINT, suggesting that as speech perception scores improved, ratings on the HAPI improved also (Mendel, 2007). While such findings are of interest to rehabilitation audiologists, further investigation is warranted before such measures are relied upon during HA evaluation. Thus, while fixed SNR measures in the aided and unaided conditions are of value to the audiological test battery, the use of these measures in demonstrating HA benefit remains a controversial issue.

An alternative method in gathering information regarding a listener's hearing ability in the presence of background noise is through the use of adaptive SNR measures. Adaptive SNR tests measure the speech-to-noise ratio as the intensity of either the speech or the noise is varied depending on the response given by the listener (Taylor, 2003). Similar to fixed SNR methods, adaptive SNR tests can be conducted in both the aided and the unaided condition (Taylor, 2003).

Two commercially available measures that employ adaptive SNR methods are the HINT (Nilsson et al., 1994) and the QuickSIN (Killion et al., 2004). The HINT measure utilises sentence stimuli that are presented alongside competing speech-shaped background noise (Nilsson et al., 1994). The noise is presented at a fixed intensity while the sentence stimuli varies in 2 dB steps (Nilsson et al., 1994). Listeners are required to identify all key words in a sentence to determine a correct response (Nilsson et al., 1994). The QuickSIN, a faster alternative to this measure, requires listeners to identify five key words from each sentence presented alongside a four-talker babble (Killion et al., 2004). The intensity at which the sentences are presented remains fixed while the background noise is varied to alter the SNR in 5 dB steps (Taylor, 2003). Unlike the HINT, the QuickSIN (Killion et al., 2004) is scored at the word-level, thus awarding a correct response for each word recalled correctly (Taylor, 2003).

Sentence-style speech-in-noise measures such as these are thought to be a valuable clinical resource as information regarding the individual's SNR loss, a facet of an individual's hearing that is not quantifiable from the audiogram, can be attained (Wilson, 2003). "SNR loss" refers to the increase in the SNR required by a listener in order to correctly identify 50% of the words in a sentence (Killion et al., 2004). It has been postulated that SNR loss may account for the variability between clients' perceived deficits despite the similar HI depicted on the audiogram (Killion et al., 2004). Therefore, it is believed that obtaining information regarding such deficits may aid a clinician's ability to recommend technology appropriate for a given client's needs (Killion et al., 2004). However, the role of SNR loss in audiologic rehabilitation is yet to be well established in empirical research, necessitating caution in the interpretation of such statements.

1.5 Selecting a Stimulus: Word versus Sentence Stimuli

When selecting a speech recognition measure for clinical use, a key consideration involves the speech material employed by a given test (Wilson, 2003). As discussed, despite the frequent use of word recognition measures in speech audiometry there are a number of drawbacks that require consideration in the selection of such tools for clinical purposes (Bosman & Smoorenburg, 1995; Ozimek, Kutzner, Sęk, & Wicher, 2009). Given that everyday communication generally involves listening, and responding, to spoken sentences, it has been suggested that employing sentence stimuli in speech audiometry may provide a closer approximation to an individual's communication deficits than is possible through word recognition tests (Cox, Alexander and Gilmore, 1992; Hochmuth et al., 2012; Killion et al., 2004). The differences between recognition of a sentence and a single word, generally relates to the context that is provided by sentence material (Ozimek et al., 2009). Contextual cues enable a listener to deduce any words in an utterance that were unintelligible (Ozimek et al., 2009). This process is exploited in everyday communication, thus supporting the rationale for implementing sentence stimuli in speech audiometry. It has also been stated that the use of sentence stimuli may increase the validity of the measure based on the greater dynamic range achieved than is possible with monosyllabic words, thus enabling a greater index of an individual's communication difficulties to be captured (Killion et al., 2004). Finally, research has indicated that sentence tests generally yield steeper intelligibility functions, as compared to tests employing isolated words, therefore resulting a more accurate measure of SRT (Bosman & Smoorenburg, 1995; Versfeld, Daalder, Festen & Houtgast, 2000). Thus, research surrounding this area almost universally supports the application of sentence recognition tests in speech audiometry due to the more comprehensive representation of an individual's deficits obtained which is thought to be of value during the rehabilitation process (Dietz et al., 2014).

There are a number of commercially available measures that aim to address the concerns associated with word recognition measures outlined by employing sentences as the target stimuli (Wilson et al., 2007a). For listeners however, identifying sentence stimuli, particularly when presented alongside masking noise, relies on many factors beyond recognition (Wilson, 2003). A wealth of literature has described the additional cognitive load associated with sentence recognition (Cervera, Soler, Dasi & Ruiz, 2009; McArdle, Wilson & Burks, 2005; Wilson et al., 2007a). It is therefore plausible that consideration of a listener's working memory ability before implementing a sentence recognition measure, as part of a diagnostic assessment, may be worthwhile. Investigating a listener's ability to complete the task may reduce the likelihood of reduced working memory capacity from counteracting the improved validity of the results expected from the use of sentence material (Craik, 1994; Kramer, Zekveld & Houtgast, 2009; McArdle et al., 2005; Wilson et al., 2007a). Thus, while there are notable benefits of incorporating sentence speech recognition tests into the diagnostic test battery, consideration of the appropriateness of a measure for a given individual is required (Wilson et al., 2007a).

1.6 Sentence Measures

A wide variety of sentence-based speech measures are available, which can be distinguished into two main categories. The first, referred to as "Plomp-type" tests (Nilsson et al., 1993; Plomp & Mimpen, 1979), employ meaningful stimulus sentences that represent natural everyday speech (Dietz et al., 2014). For example (from list 1 of the HINT; Nilsson et al., 1994, p. 1095):

"A boy fell from the window"

A commercially available example of this type of measure is the HINT (Nilsson et al, 1994). This test adaptively measures a listener's sentence speech recognition threshold (sSRT) through lists of 10 phonemically balanced sentences (Nilsson et al., 1994). Since the use of such measures was established, the HINT has been developed for other languages and dialects of English, including Cantonese (Wong & Soli, 2005), Swedish (Hällgren, Larsby & Arlinger, 2006) and NZ English (Hope, 2010). Despite such popularity, research has revealed Plomp-type tests to show a high degree of redundancy, having implications regarding its use in settings whereby frequent re-testing is required (Dietz et al., 2014). The use of sentences thought to occur in everyday speech, however, is thought to make such measures appropriate for use in diagnostics.

The second distinguishable sentence measure is the matrix sentence test (MST), originally developed by Hagerman (1982) for the Swedish language. The objective in developing this alternative sentence measure was to create a speech intelligibility test that was fast, reliable and able to be used in HA evaluation (Hagerman, 1982). MSTs are comprised of semantically unpredictable sentences of equal difficulty, from which performance can be evaluated at either the word or sentence level (Hagerman, 1982; Ozimek et al., 2010). The sentences were generated from a 50-word base matrix of five columns containing 10 names, 10 verbs, 10 numerals, 10 adjectives and 10 nouns respectively (Ozimek et al., 2010). Thus, the target sentences were created by selecting one word from each column to comprise five-word sentences. For example (translated into English from the original Swedish version; Hagerman, 1982, p. 80):

“Karin gave two old buttons”

Due to the identical structure of each of the sentences, new lists can be generated by randomly selecting words from each column, therefore making it possible to create a total of 100,000 different sentences from the original lists (Hagerman, 1982). Therefore, based on this virtually unlimited number of sentences, the MST is considered to be a useful tool in research and rehabilitation whereby repeated administration may be required (Dietz et al., 2014). Since its development the MST has gained international attention, leading to the development of a number of versions designed for different languages, including German (Wagener, Brand & Kollmeier, 1999; Wagener et al., 2014), Danish (DANTALE II; Wagener, Josvassen & Ardenkjoer, 2003), British English (Hall, 2006), Norwegian (Øygarden, 2009), Polish (Ozimek et al., 2010), Spanish (Hochmuth et al., 2012), French (Jansen et al., 2012), Russian (Warzybok et al., 2015), Dutch (Houben et al., 2014), Finnish (Dietz et al., 2014), and Italian (Puglisi et al., 2014).

1.7 The Development of the University of Canterbury Auditory Visual Matrix Sentence Test

1.7.1 **Overview**

As stated, monosyllabic word stimuli presented in quiet are typically implemented in speech audiometry conducted in clinical practice in NZ (Orchik et al., 1979). Therefore in order to progress the audiological test battery to include measures that more closely capture the deficits faced in real world listening environments, and to correspond with international trends, the need for a MST in NZ English was recognised. Originally created by Trounson and O'Beirne (O'Beirne et al., 2015; Trounson, 2012), the development of the UCAMST aimed to fulfil these requirements.

Despite its availability, employing the British English version (Hall, 2006) in NZ would compromise its validity, due to the notable differences in phonology between these dialects. NZ English is widely recognised for the differences in vowel formant structure and the raised place of production of vowels, compared to other dialects of English (Gordon et al., 2004; Maclagan & Hay, 2007). Such differences explain the variation in the pronunciation of a number of words across dialects of

English, of which may lead to errors in identification (Trounson, 2012). For example, use of the word "desks" in the British English MST (Hall, 2006) was deemed unsuitable for the UCAMST due to the possibility that NZ listeners may confuse it for the word "disks" (Trounson, 2012). This phenomenon has been described in the literature, suggesting that speech recognition, particularly in adverse conditions such as in the presence of competing background noise, can be significantly impacted when listening to a "non-native" speaker (Hochmuth et al., 2012; van Wijngaarden, Steeneken & Houtgast, 2002; Zokoll et al., 2013).

Based on this premise it was therefore necessary to develop a MST tailored for use with NZ listeners that, although based on the British English version (Hall, 2006), differed to account for the differences in the phonology of NZ English described (Trounson, 2012). Figure 3 depicts the base matrix for the UCAMST and identifies the words that replaced those in the British English version (Hall, 2006).

Name	Verb	Quantity	Adjective	Object
Amy	bought	two	big	bikes
David	gives	three	cheap	books
Hannah	got	four	dark	coats
Kathy	has	six	good	hats
Oscar	kept	eight	green	mugs
Peter	likes	nine	large	ships
Rachel	sees	ten	new	shirts
Sophie	sold	twelve	old	shoes
Thomas	wants	some	red	spoons
William	wins	those	small	toys

Figure 3. Base matrix of the UCAMST. *Retrieved from Trounson (2012, p. 24).*
Note. Dashed boxes indicate the words that were replaced for the UCAMST.

Such replacements had two main goals – first, as alluded to, to avoid vowels that may cause confusion for NZ listeners during open set testing, and second, to achieve a balance in syllables, phonemes and gender across test lists (Trounson, 2012). The changes associated with the second goal sought to achieve a balanced number of syllables within word groups, to match the language-specific phoneme distribution, and to have an equal number of names associated with each gender included in the base matrix (Hochmuth et al., 2012; Trounson, 2012). Table 1 outlines the rationale corresponding to each of the changes made to the British English MST (Hall, 2006).

Table 1.

Rationale for the changes made to the British English MST (Hall, 2006) in the development of the UCAMST. Information obtained from Trounson (2012, p. 25)

Type	Word that appears in the British English Matrix (Hall, 2006)	UCAMST changes	Rationale
Name	Alan	Amy	To achieve gender and phonemic balance
	Barry	David	To achieve phonemic balance
	Lucy	Oscar	To achieve gender and phonemic balance
	Steven	Sophie	To achieve gender and phonemic balance
	Nina	William	To achieve gender and phonemic balance
Number	Five	Those	Since “five” contains the same vowel as "nine”
Adjective	Pink	Good	To avoid confusion with the word “punk”
	Thin	New	To achieve phonemic balance
Object	Beds	Bikes	To avoid confusion with the word “bids”
	Chairs	Books	To avoid confusion with the word “cheers”
	Desks	Coats	To avoid confusion with the word “disks”
	Rings	Hats	To avoid confusion with the word “rungs”
	Tins	Skirts	To avoid confusion with the word “tens”

1.7.2 The UCAMST Auditory-Visual Component: The Rationale

When spoken discourse is encountered in everyday life listeners can typically both see and hear the speaker, thus enabling both auditory and visual information to be utilised in achieving successful communication (Mattheyses, Latacz & Verhelst, 2009). Exploiting the cues from both modes of listening is it believed to be particularly effective when trying to listen and communicate in challenging environments, regardless of whether the listener has HI (Tye-Murray, Sommers, & Spehar, 2007a; Tye-Murray et al., 2008; Tye-Murray, Hale, Spehar, Myerson & Sommers, 2014). More specifically, research has demonstrated that combining auditory and visual speech information while listening in the presence of competing background noise can yield significant improvements in speech perception as compared to listening alone (Spehar, Tye-Murray & Sommers, 2008; Sumby & Pollack, 1954; Tye-Murray, Sommers, & Spehar, 2007b). Further, it is thought that as the ability to hear the signal deteriorates, the reliance on visual cues significantly increases (Tye-Murray et al., 2007b). Based on this premise, it has been suggested that assessment of each of the three listening modalities (i.e. auditory, visual and auditory-visual) when assessing an individual's speech recognition ability may be useful in diagnostics (Tye-Murray et al., 2007b).

In accordance with such evidence the UCAMST was designed to incorporate three presentation modes – auditory, visual and auditory-visual. (Trounson, 2012). The ability to select the modality through which the stimulus is presented was thought to enable customisation of the test procedure in order to complement the goals of the assessment. For example, while testing in the auditory-alone condition may provide an indication of an individual's ability to exploit lip-reading cues, testing in the auditory-visual condition was thought to provide an index of the individual's ability to integrate information received from both modalities. Therefore, the aim of introducing the visual component to the MST was to enable a more accurate measure of a listener's real-world experiences to be obtained (Trounson, 2012). Subsequently, such information was expected to be conducive in providing rehabilitative recommendations, as the specific deficits that contribute to communication difficulties may be better conceptualised (Tye-Murray et al., 2007b).

1.7.3 Recording and Editing the UCAMST Sentences

As indicated, the UCAMST sentences were formulated through methodology identical to that employed in the development of previously published MSTs. Thus, sentences were generated based on the typical matrix format, whereby each five-word sentence contained a name, a verb, a number, an adjective and an object. The method used to record the sentences was derived from the development of the Danish MST (Wagener et al., 2003). In that version, 100 sentences were recorded in a way that allowed all of the words in a given column to be recorded in conjunction with all of the words in the subsequent column (Wagener et al., 2003). The reader is referred to Figure 4 whereby this recording technique is displayed for Index 0 (translated to English from the Danish MST; Wagener et al., 2003). This procedure was repeated for each of the remaining indices.

Index	*Name*	*Verb*	*Numeral*	*Adjective*	*Object*
0	Anders	owns	ten	old	jackets
1	Birgit	had	five	red	boxes
2	Ingrid	sees	seven	nice	rings
3	Ulla	bought	three	new	flowers
4	Niels	won	six	fine	cupboards
5	Kirsten	gets	twelve	lovely	masks
6	Henning	sold	eight	beautiful	cars
7	Per	borrows	fourteen	big	houses
8	Linda	chose	nine	white	presents
9	Michael	finds	twenty	funny	plants

Figure 4. Sentence recording technique utilised in the development of the test sentences for the Danish MST (English Translation; *Wagener et al. (2003, p. 13*). *Reproduced with permission. Copyright (2016) by Taylor and Francis.*

This procedure was applied in the recording of the UCAMST sentences because it accounted for coarticulation, enabling the files to be cut in a manner that preserves the natural properties of the sentences during the editing phase (Wagener et al., 2003). This recording procedure differed from the original Swedish version (Hagerman, 1982) in that only the base list sentences were recorded, without accounting for the transitions between words (Wagener et al., 2003). This method demonstrated the importance of considering coarticulation during the recording phase through the less natural-sounding test sentences produced (Wagener et al., 2003). Thus, implementing the recording methodology employed in the construction of the Danish MST (Wagener et al., 2003) was advisable for the development of the UCAMST in order to achieve smooth transitions between the words in a given sentence. Following the recording process the 400 word fragments obtained were then available to be combined to generate 100,000 unique sentences.

Despite the vast improvements in the quality of the final sentences obtained when employing such a technique, previous research has revealed that unnatural sounding final sentences can remain (Hochmuth et al., 2012; Houben et al., 2014). Where this occurred the affected sentences were removed from the final sentence lists of the UCAMST. In addition to this, however, the UCAMST had the challenge of ensuring that both the audio and the visual components appeared natural to the viewer/listener. A marked jerk (termed, and henceforth referred to as, a "judder") was evident in the visual component, where a mismatch occurred between the actress' head position between fragment transitions. Despite employing a number of precautions in order to avoid an unnatural appearance in the visual component of the stimuli, a large proportion of the synthesised sentences had a noticeable judder, thus warranting further investigation.

1.7.4 **Selecting the Sentence Stimuli**

To ensure that the final sentences were appropriate for use in both the auditory and visual conditions, McClelland (2014) conducted a study to evaluate the noticeability of the judders present. That study employed listeners with NH to subjectively rate the noticeability of the judder in sentences with and without (i.e. control condition) a present judder (McClelland, 2014). Judder ratings were selected on a continuum from 0 (no noticeable judder) to 10 (highly noticeable judder) (McClelland, 2014). Paired t-tests were performed to make multiple comparisons across the sentences (McClelland, 2014). The final sentence repertoire was comprised of the control sentences and, of the sentences that contained a judder, those that were rated to have the least noticeable judder were included (McClelland et al., 2014). This method ensured that the final pool of sentences for testing in the visual conditions was large enough (McClelland, 2014).

1.7.5 Generating the Masking Noise

Two types of masking noise were produced for the UCAMST – constant-speech-shaped noise and six-talker babble (herein referred to as "constant noise" and "babble noise", respectively). The constant noise was generated by randomly superimposing the audio recordings 10,000 times via an automated process. Therefore, the noise was created to have almost identical spectral components as the signal (i.e. they were spectrally-matched) (King, 2010). According to King (2010), spectral matching has important implications with regards to maintaining the SNR of the signal when presenting the stimuli via headphones or speakers, thus preserving the validity of the measure.

The babble noise was originally developed for use as part of a previous master's research study at the University of Canterbury (Spencer, 2011). In order to create this noise, six speakers (three males and three females) of NZ English were recorded reading 20, 6- to 10-word, semantically anomalous sentences (Spencer, 2011). Each of the sentences were then mixed into a single sound file to generate the babble noise employed in the UCAMST.

1.7.6 Normalisation of the UCAMST Sentences

The next stage in developing a new speech recognition measure, referred to as optimization, relates to achieving high homogeneity (i.e. equivalence) among test items (Akeroyd et al., 2015; Kollmeier et al., 2015). First, speech materials need to be optimized by determining the word-specific intelligibility functions for each word recorded (Akeroyd et al., 2015). Obtaining the word-specific intelligibility functions is generally achieved through administering the stimuli to approximately 10 participants with NH at fixed SNRs (Akeroyd et al., 2015). This process identifies the items that are of high and low intelligibility, to which level adjustments can be applied, where necessary, in order to produce intelligibility functions that are as similar as possible (Akeroyd et al., 2015; Kollmeier et al., 2015). Exclusion of items that do not adequately fit the word-specific intelligibility function is generally advised (Kollmeier et al., 2015).

In order to complete normalisation on the UCAMST speech materials McClelland (2014) recruited 17 participants with NH to assess 400 sentences containing each audio fragment. The stimuli were presented in both constant and babble noise at each of the following SNRs: -18.5 dB, -15 dB, -11.5 dB, and -8 dB. This aspect of McClelland's (2014) work was divided into two components – normalisation by fragment (which treated the individual recordings of words differently) and word-specific normalisation (which applied the average adjustments calculated from the individual recordings of a word to all occurrences of that word) (McClelland, 2014). The latter process is based on the assumption that it is the acoustic characteristics of the word itself, rather than the speaker's performance of it in a particular fragment, that is the dominant determinant of its intelligibility at a given SNR (McClelland, 2014). This method also has the advantage of having access to 10-times the amount of raw psychometric data than the fragment-specific process, which was believed to result in more reliable adjustments (McClelland, 2014).

1.7.7 **Fragment- and Word-Specific Normalisation**

Normalisation of the UCAMST items by fragment enabled fragment-specific intelligibility functions to be generated, thus enabling the appropriateness of the fit to be evaluated first (McClelland, 2014). Intelligibility functions were produced for each fragment by calculating the mean intelligibility (%) across SNRs (McClelland, 2014). The resulting intelligibility function was then fit to the following model, described in equation (1), adapted from Kollmeier and Wesselkamp (1997) and Wagener et al.

(2003). A conservative adjustment limit of ± 3 dB was appointed based on the methodology employed in the normalisation of previously published versions (i.e. Dietz et al., 2014; Hochmuth et al., 2012; Ozimek et al., 2012) (McClelland, 2014).

$$SI(L) = \frac{1}{A}\left(\frac{(1 + SI_{max}).(A-1)}{1 + \exp\left(\left[-4.5.\frac{slope}{100}\right].[L - L_{mid}]\right)}\right) \tag{1}$$

Note. SI = speech intelligibility; L = level; L_{mid} *= midpoint;* SI_{max} *= function ceiling; A = number of alternatives;* $\frac{1}{A}$ *= function floor.*

Normalisation by fragment was first completed for test items designed for use with the constant noise. This procedure revealed the fit of 15 fragments (i.e. 4% of the total) to be inadequate, requiring them to be removed from the final pool (McClelland, 2014). The remaining 385 fragments however, produced a pre-normalisation midpoint (L_{mid} or 50% correct point) of -10.3 dB SNR (± 2.1 dB standard deviation [SD]) (McClelland, 2014) [1]. Word-specific intelligibility functions were then fit, allowing the data to be normalised (McClelland, 2014). The L_{mid} of each word-specific intelligibility functions were adjusted to equal the mean pre-normalisation mean fragment (-10.3 dB SNR) to achieve greater overlap in the post-normalisation functions (McClelland, 2014), which in turn improves the overall slope of the test. The adjustments made from pre- to post-normalisation are depicted in Figure 5.

[1] SNR values quoted from McClelland (2014) have been corrected following the recalibration procedure described in section 2.3.1.

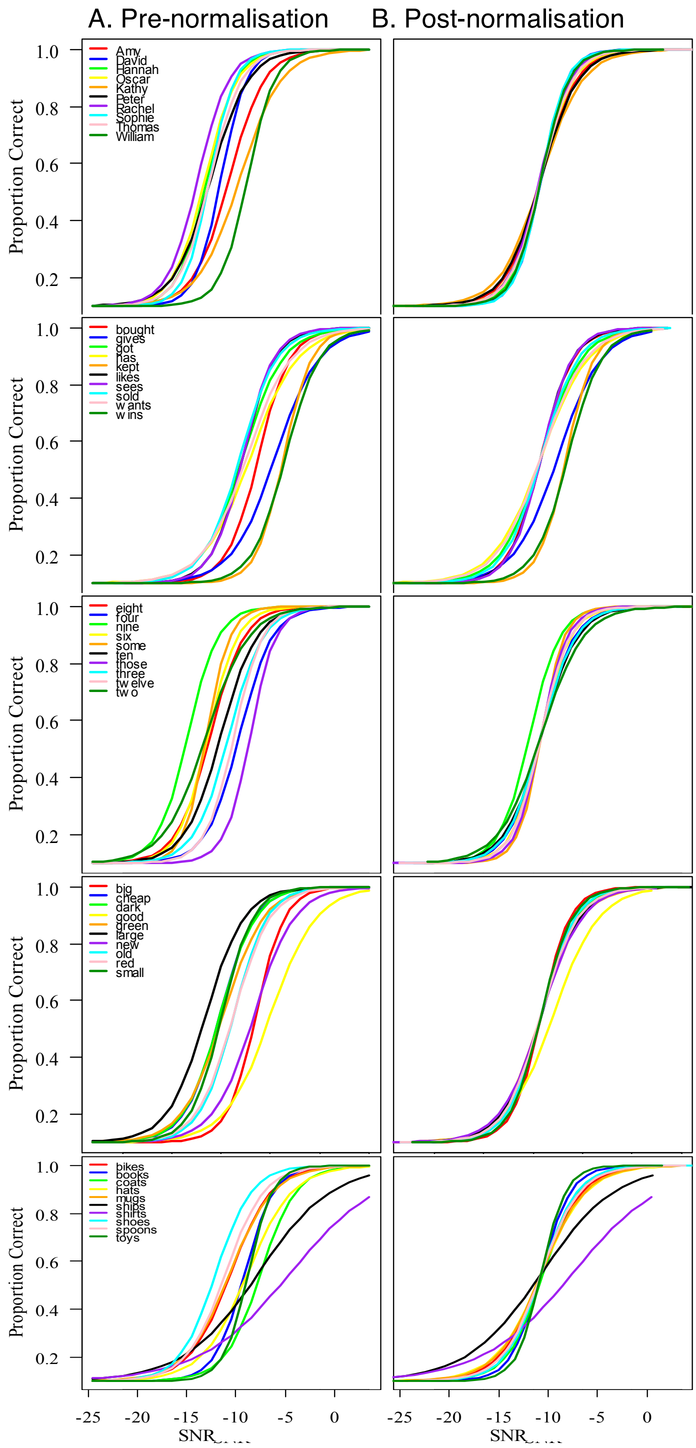

Figure 5. Post-recalibration pre-normalisation (A) and predicted post-normalisation (B) word-specific intelligibility functions for the constant noise condition. *Adapted from McClelland (2014, p. 82).*

As illustrated in Figure 5 the adjustments resulted in the aligning of the L_{mid} and, consequently, a greater overlap in the post-normalisation functions, as compared to the pre-normalisation functions (McClelland, 2014). It is noteworthy that the words “shirts” and “ships” required adjustments in excess of the limit, in order to achieve appropriate alignment with the other functions (McClelland, 2014). Despite this however, the mean word-specific L_{mid} for this condition was predicted to be -10.1 dB SNR ± 0.8 dB (SD), thus denoting a 1.6 dB decrease in the SD of the L_{mid} measures for words designed to be used in the constant noise condition (McClelland, 2014).

As stated, the test items designed for use with the babble noise were then normalised utilising the procedure described above, with fragment-specific normalisation preceding word-specific normalisation. Fragment-specific normalisation in this condition resulted in 47 fragments being discarded due to the inability to fit the model (McClelland, 2014). The L_{mid} across the remaining 353 fragments was -11.0 dB SNR (± 2.9 dB [SD]), indicating that the UCAMST test items presented amongst babble noise were easier to detect than those in the constant noise condition (McClelland, 2014). Word-specific intelligibility functions were then fit in order to normalise the word-specific functions. Examination of the L_{mid} for each word-specific function revealed that 20 words (i.e. 41% of total) required adjustments that exceeded the limit (McClelland, 2014). The post-normalisation functions for each word position are illustrated in Figure 6.

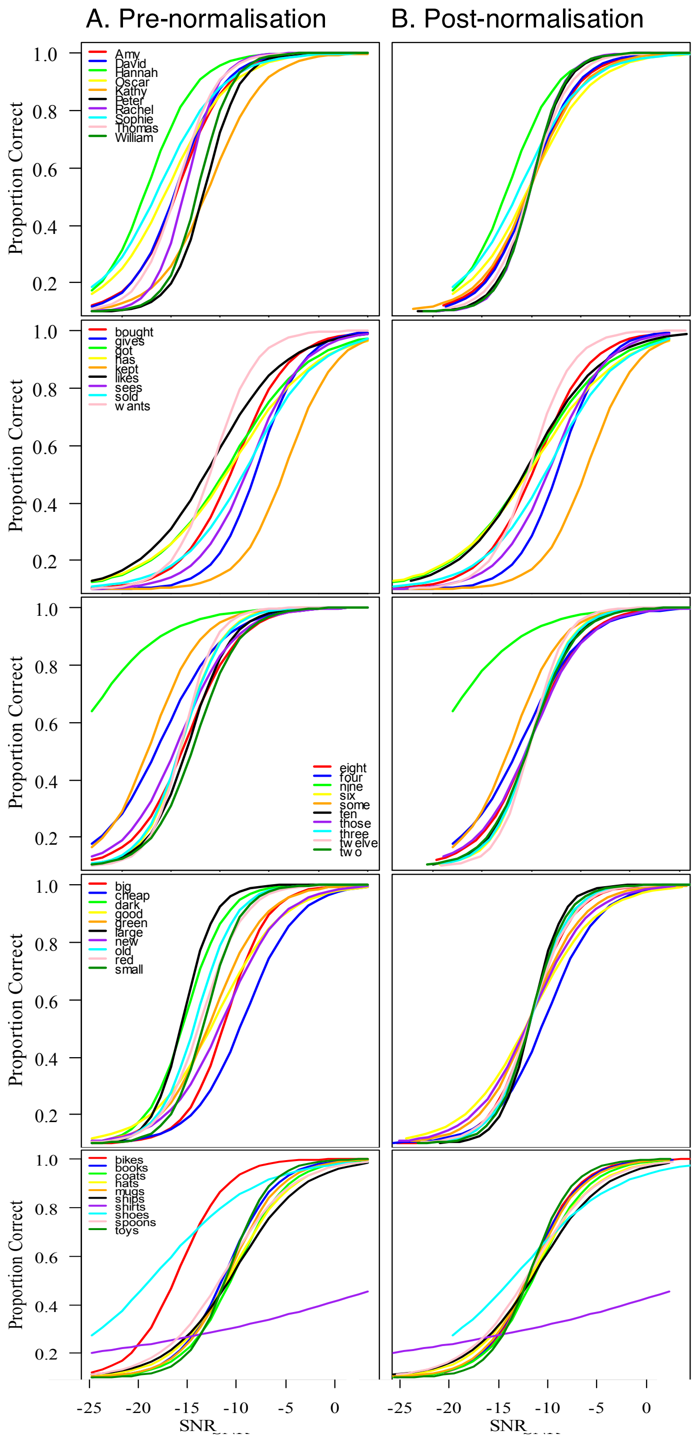

Figure 6. Post-recalibration pre-normalisation (A) and predicted post-normalisation (B) word-specific intelligibility functions for the babble noise condition. *Adapted from McClelland (2014, p. 87).*

The result of the normalisation process for the UCAMST test items revealed less overlap between the predicted post-normalisation functions of test items utilised with the babble noise than those used with the constant noise (refer to Figures 5 & 6 for comparison) (McClelland, 2014). McClelland (2014) asserted that this difference resulted from the larger quantity of words requiring adjustments in excess of the limits in the babble condition, as compared to the constant noise condition. Despite such disproportionate changes, the final result of the normalisation process for the babble test items revealed a mean post-normalisation L_{mid} of -11.0 dB SNR (±1.9 dB [SD]), denoting a reduction of 1.7 dB in the SD of word-specific L_{mid} measures for words in this condition (McClelland, 2014).

1.8 Evaluating the Normalisation

In order to evaluate the normalisation, the slope of the test-specific ($s50_{test}$) function needs to be examined. This process enables the equivalence of the test lists to be assessed, thus providing confidence that, regardless which list is administered, the same SRT and the same slope of the intelligibility function is likely to be obtained (Akeroyd et al., 2015; Kollmeier et al., 2015). The test-specific intelligibility function equates to the convolution of the mean word-specific function and the SD of the SRTs, described by equation (2), adapted from Hochmuth et al. (2012):

$$s50_{test} = \frac{s_{word}}{\sqrt{1 + \frac{16 s_{word}^2 \times \sigma_{Lmid}^2}{\left(\ln\left(2e^{\frac{1}{2}} - 1 + 2e^{\frac{1}{4}}\right)\right)^2}}} \tag{2}$$

Note. *$s50_{test}$ = test-specific speech recognition curve; s_{word} = slope of the word-specific intelligibility function; σ = standard deviation of word-specific L_{mid} measures.*

Due to stringent time constraints, the evaluation of the normalisation process was unable to be conducted for the UCAMST in McClelland's (2014) work and is therefore the foundation of the current research. Preliminary evaluation of the tests was conducted in McClelland's (2014) work based on predicted post-normalisation values, however. This process enabled a predicted outcome of the $s50_{test}$ for both the constant and babble noise conditions to be obtained (McClelland, 2014). A summary of the pre-normalisation measurements and post-normalisation predictions of the mean L_{mid}, mean word-specific slope (s_{word}), and the $s50_{test}$ values are provided in Table 2 below.

Table 2.

Predicted outcomes from the normalisation process for word stimuli *(updated from McClelland (2014) following recalibration, detailed in section 2.3.1).*

	Constant noise		Babble noise*	
	Pre-normalisation measurement	Post-normalisation prediction	Pre-normalisation measurement	Post-normalisation prediction
Mean L_{mid} (dB SNR)	-9.77 ± 2.41	-10.11 ± 0.75	-10.71 ± 3.26	-10.95 ± 1.30
Mean s_{word}	14.38%	14.38%	10.26%	10.26%
$s50_{test}$	10.80%	13.90%	7.80%	9.70%

*Note. * denotes the removal of "shirts" and "wins" from the lists.*

1.9 Selecting the Presentation Mode

In order to discuss the current study it is important to first address another key consideration when developing a MST – the mode through which the stimuli will be presented. For MSTs, speech materials can be presented in either the closed-set mode, whereby the response alternatives are selected from a visible word matrix, or the open-set mode, where no such cues are provided and the listener verbally recalls the items recognised (Hochmuth et al., 2012). A key advantage of utilising the closed-set mode lies in the ability to record a listener's performance without the need for an instructor (i.e. audiologist or researcher) to be involved in the test procedure (Hochmuth et al., 2012). This is made possible through the use of a touch-sensitive monitor displaying the response alternatives and instructing the listener to select the words that were perceived. Previously published MSTs have employed both open- (Dietz et al., 2014; Ozimek et al., 2010; Wagener et al., 2003) and closed-set (Houben et al., 2014) modes.

The effects of the presentation mode on performance have been highlighted in the literature to date, however the findings are somewhat ambiguous between studies. Ozimek et al. (2010) investigated this phenomenon and found no significant differences between the SRTs of those tested in the open- and closed-set modes. Conversely, Hochmuth et al. (2012) found a significant difference between the SRTs obtained using each presentation mode ($F(1, 41) = 22.30$, $p < 0.001$). However, investigation of such differences revealed that the number of training trials provided across the studies may have influenced the findings (Hochmuth et al., 2012). It is possible that the listeners' involved in Ozimek et al.'s (2010) research were more familiar with the test materials, due to the increased number of training sentences performed, thus improving overall performance. Based on these findings, preliminary research has suggested the importance of training in preserving the validity of the results, and consequently the equivalence of listener performance between the two test conditions.

1.10 Study Rationale

The current research sought to continue the work of McClelland (2014) through evaluating the normalisation of the stimulus lists employed in the UCAMST. Evaluation is the necessary next stage in developing a MST as the reliability and sensitivity of the measure in estimating SRTs will be determined. Therefore the current research is essential in progressing the UCAMST toward clinical use as part of the University of Canterbury Adaptive Speech Test (UCAST; O'Beirne, McGaffin & Rickard, 2012) platform. The aim of the UCAST is to comprise a suite of audiological tests including the NZHINT (Hope, 2010) and the NZ Digit Triplet Test (NZDTT; King, 2011) available for clinical and research use (O'Beirne et al., 2012). Thus, once the final stages of development have been completed for the UCAMST it will be able to be integrated into this battery of tests.

It is noteworthy that the normalisation process has only been completed for the auditory-alone condition to date, based on the preliminary findings of research conducted with the Malay version of the UCAMST (Jamaluddin & O'Beirne, 2015). This research revealed that presenting sentences at poor SNRs in the auditory-visual condition was equivalent to testing in the visual-alone condition, as listeners were reliant on the visual cues provided in these conditions. This therefore created difficulty in obtaining a psychometric function for the auditory-visual condition, hence the need to exclude the visual components during the normalisation process of the UCAMST (McClelland, 2014).

1.11 Evaluation of the UCAMST

The evaluation of the UCAMST followed the guidelines provided by Akeroyd et al. (2015) and the methods utilised by previously published MSTs. Such methodology, and the results obtained, which will be reviewed in the following sections.

1.11.1 **The Danish MST (Wagener et al., 2003)**

Sixty adult listeners with NH were employed in order to evaluate the normalisation of the DANTALE II (Wagener et al., 2003). Two SNRs (-10 dB SNR and -6 dB SNR) were selected for the process based on the estimates from the optimisation process of the corresponding intelligibility above and below 50% (Wagener et al., 2003). Participants were divided into two groups whereby half performed half of the lists at the lower SNR first followed by the remaining sentences at the higher SNR, and vice versa for the other group (Wagener et al., 2003). All test lists were presented amongst constant noise presented at 65 dB SPL. Test lists were presented according to an adaptive procedure whereby the listener's response to the preceding stimulus determined the presentation of the following trial (i.e. if the listener responded incorrectly the stimulus level was increased and vice versa for a correct response) (Brand and Kollmeier, 2002; Levitt, 1971). According to an adaptive procedure, across the course of the test the presentation level begins to converge around the listener's SRT, allowing the presentation levels to be averaged in order to reveal the final SRT (Levitt, 1971). Thus, the objective behind implementing this technique for the evaluation phase was to be efficient in obtaining estimates of the listeners' SRTs (Wagener et al., 2003).

Results of the evaluation procedure revealed a mean list-specific SRT of -8.38 dB SNR, with a SD of 0.16 dB SNR between test lists, and an accompanying slope of 12.6 %/dB (± 0.8 dB) (Wagener et al., 2003). Further, examination of the data via a single analysis of variance (ANOVA) identified no significant difference between the intelligibility of the test lists at each of the SNRs – $F = 0.80$ at -10 dB SNR and $F = 1.26$ at -6 dB SNR (Wagener et al., 2003). This finding was concluded to suggest test-retest reliability across estimates of SRT of approximately 1 dB when administering 20 sentences (Wagener et al., 2003).

1.11.2 **The Polish MST (Ozimek et al., 2010)**

Thirty listeners with NH were recruited for the evaluation of the polish MST (Ozimek et al., 2010). Test stimuli were presented alternately at two SNRs: -11 dB

SNR and -7 dB SNR in an attempt to approximate 20% and 80% intelligibility and therefore the so-called “pair of compromise” (Ozimek et al., 2010). The pair of compromise, thought to be located at the SNRs at which 19% and 81% intelligibility is obtained, has been postulated to yield highly accurate simultaneous measurements of SRT and $s50_{test}$ in an efficient manner, thus explaining the use of such methods in evaluation procedures (Brand & Kollmeier, 2002; Ozimek et al., 2010). The procedure was designed in a way whereby the even lists (i.e. 2nd, 4th, …, 10th) were presented to half of the listeners at the lower SNR and the odd lists (i.e. 1st, 3rd, …, 9th) were presented at the higher SNR and vice versa for the remaining participants (Ozimek et al., 2010). Each participant was required to listen to 10 lists of 10 sentences at the two SNRs. All stimuli were presented amongst babble noise presented at a constant intensity of 65 dB SPL.

In order to analyse the evaluation procedure, a system of two logistic functions, depicted in equations (3) and (4), were solved for each of the SNRs to obtain list-specific SRT and $_{s}50$ values (Ozimek et al., 2010, p. 449).

$$P(-11) = \frac{100}{1 + exp^{(-\frac{(1-11-SRTl)}{S50l})}} \tag{3}$$

$$P(-7) = \frac{100}{1 + exp^{(-\frac{(1-7-SRTl)}{S50l})}} \tag{4}$$

Note. SRT_l *= list-specific SRT;* $S50_l$ *= list-specific slope at the SRT in equations (4) and (5).*

Results revealed the mean SRT and slope to be -9.6 dB and 17.1%/dB, respectively, therefore inferring that the adjustments made throughout the normalisation procedure resulted in better equalisation of intelligibility across test lists (Ozimek et al., 2010).

1.11.3 **The Spanish MST (Hochmuth et al., 2012)**

The evaluation procedure was conducted independently for open and closed set test conditions for the Spanish MST (Hochmuth et al., 2012). The open set procedure will be discussed first (Hochmuth et al., 2012). Test stimuli were presented amongst constant noise set at 65dB SPL, at fixed SNRs of -4 dB, -5 dB and -9 dB for this condition in order to efficiently estimate the points of 80%, 50% and 20% speech recognition (Hochmuth et al., 2012). A total of 33 adult listeners with NH were recruited in order to evaluate the open set condition. Results revealed the SD of the SRT to be 1.1 dB (Hochmuth et al., 2012). Further investigation via a two-way repeated-measures ANOVA (RM-ANOVA) revealed a significant difference on the test list factor (F (11, 341) = 4.624, $p < 0.001$; Hochmuth et al., 2012). Pairwise comparisons applying a Bonferroni correction revealed significant differences between list 1 and 6 ($p = 0.013$) and list 3 and 2, 4, 6, and 7 ($p = 0.015$, $p = 0.011$, $p = 0.012$, and $p = 0.005$, respectively; Hochmuth et al., 2012). Based on these analyses lists 1 and 3 were excluded from the final test resulting in a reduction in the SD across test lists from 0.2 dB to 0.13 dB (Hochmuth et al., 2012).

Following the evaluation of the open set condition, the closed set condition was investigated. A total of 10 adult participants with NH were recruited for the evaluation of the test stimuli for this condition (Hochmuth et al., 2012). SNRs of -4 dB and -9 dB were selected based on expected recognition rates of 80% and 20% respectively (Hochmuth et al., 2012). The logistic model described by equation (5) was fitted to the data, excluding lists 1 and 3, based on the open set findings, and resulted in an SRT of -7.7 dB SNR and a slope of 14 %/dB (Hochmuth et al., 2012, p. 538).

$$SR\ (SNR) = \frac{100}{1 + e^{(-4_s 50\ (SNR - SRT)}} \tag{5}$$

Note. SR = speech recognition in percentage; $_s50$ = the slope at the SRT; SNR = Signal-to-Noise Ratio; SRT = Speech Recognition Threshold.

The authors concluded, based on the outlined findings, that the assumption can be made that the test lists designed for both conditions provide consistent recognition rates and can therefore be used interchangeably (Hochmuth et al., 2012).

1.11.4 The French MST (Jansen et al., 2012)

Twenty participants with NH were recruited for the evaluation phase of the French MST (Jansen et al., 2012). Prior to completing the test procedure, participants completed as training phase consisting of six double lists (i.e. 12 lists of 10 sentences) (Jansen et al., 2012). This notable increase in the number of practice lists completed by participants, compared to those utilised in the evaluation of other MSTs (i.e. the Dutch MST; Houben et al., 2012), was employed to enable the training effect associated with this version to be evaluated concurrently (Jansen et al., 2012). Following the practice phase, the test procedure commenced. The stimulus sentences were presented alongside stationary speech noise at a fixed level of 65dB SPL at the following SNRs – -8.0 dB, -6.5 dB, -5.0 dB, and -3.5 dB – in order to yield intelligibility scores above and below 50% (Jansen et al., 2012).

The results of this procedure yielded an average SRT of -6.0 dB SNR (± 0.6 dB) and an average slope at the SRT of 14.0%/dB (Jansen et al., 2012). In addition to this, the list-specific SRT was determined by pooling all of the data collected together. The SD of the SRTs across each of the lists was 0.1 dB, indicating minimal deviation between the estimates of SRT that would be obtained across the test lists (Jansen et al., 2012). Accordingly, it was concluded that the stimulus lists incorporated into the French MST (Jansen et al., 2012) were adequately equivalent and were therefore appropriate in providing reliable estimates of SRT.

1.11.5 **The Finnish MST (Dietz et al., 2014)**

In order to evaluate the Finnish MST, Dietz et al. (2014) recruited 21 adult participants with NH. The measurements were performed at constant SNRs of -12.5 dB SNR, -10.5 dB SNR, and -8.5 dB SNR in order to approximate the points whereby participants will score 20%, 50% and 80% respectively (Dietz et al., 2014). The test lists were randomised and were presented amongst constant noise set at 65 dB SPL (Dietz et al., 2014).

Utilising this procedure enabled list-specific speech recognition functions to be produced and inter-individual differences between test participants to be evaluated (Dietz et al., 2014). Results revealed the slope of the lists and the mean SRT to be 16.7 dB SNR (± 1.2 dB SNR) and -10.1 dB SNR (± 0.1 dB SNR) respectively (Dietz et al., 2014). The mean SRT and slope revealed for participants were -10.1 (± 0.7 dB SNR) and 17.5 dB SNR (± 2.2 dB SNR) respectively (Dietz et al., 2014). In combination, such results indicate the test lists for this measure to be interchangeable.

1.11.6 **The Dutch MST (Houben et al., 2014)**

Evaluation of the Dutch MST (Houben et al., 2014) was conducted across three centres located in Belgium, Rotterdam and Amsterdam. Each centre recruited 15 adults with NH in order to assess the equivalence of the stimuli incorporated into the measure (Houben et al., 2014). Prior to data collection, each participant was familiarised with the test format through completion of two practice lists (Houben et al., 2014). Test stimuli were presented at fixed SNRs of -5 dB, -7dB and -9 dB amongst stationery speech noise (Houben et al., 2014).

To investigate the comparability of the data across the three centres a logistic regression model that described intelligibility as a function of SNR was applied. Given that the Dutch MST implemented a closed set test format, of which produces the probability that correct recognition is due to chance 10% of the time, application of this model (depicted in equation 6) was important as it accounted for this effect (Houben et al., 2014, p.763).

$$\log\left(\frac{(p-a)}{(1-p)}\right) \tag{6}$$

Note. p represents the probability that the sentence is correctly repeated by the listener.

The ANOVA conducted on the data revealed no differences between the SRT and slope across centres – $F(2, 42) = 0.04$, $p = 1$ and $F(2, 42) = 0.9$, $p = 0.4$, respectively (Houben et al., 2014). Following this the intelligibility functions were fit, revealing the average list-specific SRT across the stimulus lists to be -8.4 dB SNR (± 0.2 dB SNR) and the average slope to be 10.2 %/dB (± 0.9 %/dB) (Houben et al., 2014). Based on these findings in combination, the authors concluded that the stimulus lists included in the Dutch MST (Houben et al., 2014) were homogenous and that they were therefore appropriate for use in both the Netherlands and in Belgium.

1.11.7 **The Italian MST (Puglisi et al., 2014)**

Eleven adult native speakers of Italian with NH were recruited for the evaluation phase of the Italian MST (Puglisi et al., 2014). The procedure employed for this phase of the measure's development required participants to complete six double lists at fixed SNRs of -4.5 dB, -7 dB, and -9.5 dB, believed to correspond to recognition rates of approximately 80%, 50%, and 20%, respectively (Puglisi et al., 2014). Test materials were presented amongst spectrally-matched speech noise, developed according to the procedure implemented by Wagener et al., (1999), at a fixed an intensity of 65 dB SPL (Puglisi et al., 2014).

In order to determine list equivalence, the average intelligibility scores were averaged across participants and fit to the logistic function depicted in equation (5) (Puglisi et al., 2014). The findings revealed a mean list-specific SRT of -7.3 dB SNR (± 0.2 dB SNR) and slope of 13.3 %/dB (±1.2 %/dB) (Puglisi et al., 2014). Based on these results the researchers concluded that the Italian MST (Puglisi et al., 2014) stimulus lists to be equivalent, making it a useful assessment tool, particularly where repeated measurements are required.

1.11.8 **The Russian MST (Warzybok et al., 2015)**

Evaluation of the open set condition of the Russian MST (Warzybok et al., 2015) necessitated completion of the test procedure by 20 adult listeners with NH. A total of eight lists of 20 sentences were presented to participants at fixed SNRs thought to be located at the pair of compromise (Warzybok et al., 2015). The pair of compromise was set at -11.2 dB SPL and -8.2 dB SPL for this measure, with each list presented at each SNR in a random order (Warzybok et al., 2015). The noise level was developed according to the procedure described by Wagener et al., (1999) in order to generate a masker that matched the power spectrum of the sentences. This method was employed as it was thought that it would increase the likelihood of obtaining a steep list-specific intelligibility function, thus deeming the measure to be reliable (Warzybok et al., 2015).

The average intelligibility scores were fit to the logistic model utilised in the evaluation of both the Spanish MST (Hochmuth et al., 2012) and the French MST

(Puglisi et al., 2014), described in equation (5) (Warzybok et al., 2015). From this the mean SRT and slope of the test-specific function were revealed to be -9.5 dB SNR (± 0.2 dB SNR) and 13.8 %/dB (± 1.6 %/dB), respectively (Warzybok et al., 2015). Further, a RM-ANOVA was conducted to statistically test the equivalence of the stimulus lists, revealing no significant differences with regards to SRT and slope ($F(1, 19)= 1.03$, $p = 0.329$, Greenhouse-Geisser correction and $F(1.9, 20.51) = 1.21$, $p = 0.259$, Greenhouse-Geisser correction, respectively; Warzybok et al., 2015). Accordingly, the researchers concluded that the lists incorporated into the measure were found to be homogenous and are therefore appropriate for interchangeable use (Warzybok et al., 2015).

1.12 Aims and Hypotheses

This thesis aimed to generate test lists appropriate for use in each of the presentation modes included in the UCAMST design and evaluate the difficulty of such lists. In order to evaluate list equivalence, this study sought to answer three primary research questions:

(1) Are the stimulus lists designed for use in each condition (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise) equivalent with regards to:

   a) Slope

   b) The SNR at which SRT is estimated (herein referred to as SRT)

(2) Is there a difference between the slope and SRT of the four test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise)

(3) Are the stimulus lists designed for use in the UCAMST equivalent to previously published MSTs (Dietz et al., 2014; Houben et al., 2014; Jansen et

al., 2012; Øygarden, 2009; Ozimek et al., 2010; Puglisi et al., 2014; Wagener et al., 2003; Warzybok et al., 2015) with regards to:

a) Slope

b) SRT

When developing MSTs the methodological standards that are to be employed are stringent (Dietz et al., 2014). Therefore, when conducting the evaluation procedure, previous literature have revealed the test lists to be equivalent with regards to slope and SRT, with any minor differences attributable to language- or speaker-dependent factors (Dietz et al., 2014; Hochmuth et al., 2012). Further, evaluation of previously published MSTs also revealed evidence of equivalence across tests designed for different languages, a finding that is likely due to the analogous methodology employed by each researcher (Dietz et al., 2014; Ozimek et al., 2010). Based on the findings of previous research the following hypotheses were proposed for the current study:

For research question (1):

(1) That no significant differences would be found between the stimulus lists in the <u>closed set, constant noise</u> condition for:

a) Slope

b) SRT

(2) That no significant differences would be found between the stimulus lists in the <u>open set, constant noise</u> condition for:

a) Slope

b) SRT

(3) That no significant differences would be found between the stimulus lists in the <u>closed set, babble noise</u> condition for:

a) Slope

b) SRT

(4) That no significant differences would be found between the stimulus lists in the open set, babble noise condition for:

a) Slope

b) SRT

For research question (2):

(5a) That no significant difference would be found between the four test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise) with regards to *slope*.

(5b) That no significant difference would be found between the four test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise) with regards to *SRT*.

For research question (3):

(6) That no significant differences would be found between the stimulus lists designed for use in the UCAMST and those of previously published MSTs (Dietz et al., 2014; Houben et al., 2014; Jansen et al., 2012; Øygarden, 2009; Ozimek et al., 2010; Puglisi et al., 2014; Wagener et al., 2003; Warzybok et al., 2015) with respect to:

a) Slope

b) SRT

CHAPTER TWO:

METHODS

2.1 Overview

As described, the purpose of the current research was to determine the equivalence of the test lists designed for use in the UCAMST. To achieve this aim, a large sample of listeners with NH was required for this research. The following chapter discusses the methodology employed in the current research, including the participants recruited, instrumentation and stimuli employed, the procedure utilised and the statistical analyses applied to the data.

Prior to commencing the current research, an ethics application was submitted to the University of Canterbury Human Ethics Committee and approval was acquired on 11 May 2015 (refer to Appendix A for a copy of the approval letter received). All procedures conducted in the current research complied with those proposed in the application.

2.2 Participants

2.2.1 **Recruitment**

In order to complete the evaluation process in accordance with the procedure employed by previous researchers, a sample of 64 participants was required for the current research. This number of participants was determined to provide accurate estimates of SRT for each list through providing eight approximations of the SRT at each SNR for each list in each condition. Recruitment was primarily conducted within the University of Canterbury community (Christchurch, NZ via the circulation of advertisements and an email invitation. As shown in Appendices B.1 and B.2, respectively, these invitations briefly detailed the aims and nature of the study in addition to the inclusion criteria that eligible participants were required to meet. Participants were required to give informed consent prior to involvement in any of the current research procedures. This process ensured that all participants understood the requirements of, and risks associated with, being a research participant in the current study. The information sheets and consent forms developed for the current research are provided for reference in Appendices C.1 and C.2, respectively. Once consent was obtained, candidates were screened, via a hearing test and an interview, in order to determine whether they met the inclusion criteria of the study (outlined in Table 3).

Table 3.

Participant inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criterion
Aged 18 years (or over)	An identified HI or air-bone gap (ABG) of ≥15dB HL across the following test frequencies: 500, 1000, 2000 and 4000 Hz
NH (defined as thresholds of ≤20dB HL at octave frequencies between 250 – 8000 Hz)	
Native speaker of NZ English	

The inclusion criteria were selected based on the empirical evidence suggesting such characteristics to have a likely influence on the data collected. First, participants were required to be 18 years of age due the length of time required to complete the study, and therefore the extended period of time for which attention would need to be sustained. It has been extensively recognised in the literature that the ability to sustain attention to a task (i.e. vigilance) continues development into adolescence (Betts, Mckay, Maruff, & Anderson, 2006; Rebok et al., 1997). Further, it is thought that performance is greatly influenced by factors such as task duration and complexity (Betts et al., 2006). Therefore, given that the task involved in the current study was considered to be one of high cognitive load and required sustained attention for 60 minutes, adult participants were recruited. Furthermore, it was necessary to ensure that the methodology involved in the development of the UCAMST was in accordance with those employed in the development of previously published MSTs. Therefore, given that such research generally recruited adult participants, the current sample was also restricted to those over 18 years of age. The second inclusion criterion, listeners with NH, was employed as it was essential to ensure that a HI did not confound the data obtained (Akeroyd et al., 2015). Last, the listeners involved in the study were required to be native speakers of NZ English in order to preserve the validity of the findings. As discussed, speech intelligibility can be significantly compromised when listening to a "non-native" speaker and therefore, in order to determine the use of the UCAMST in a NZ context, native speakers of NZ English were required (van Wijngaarden et al., 2002; Zokoll et al., 2013).

The exclusion criterion was employed to ensure that no participants had HI, of any nature. The specification of an ABG of ≥15 dB HL in the exclusion criterion was established since it is possible that an ABG can be indicative of current middle ear pathology (Hussain, 2008). It is recognised that middle ear pathology can lead to either permanent or temporary shifts in the individual's hearing thresholds, thus presenting a HI on the pure tone audiogram (Hussain, 2008). Thus, excluding those who presented with a HI, of any nature, was done with the aim of strengthening the validity of the current research results.

All participants, including those identified to have a HI during the initial screening phase and therefore were not eligible to complete the full procedure, were offered an inducement of a $10 Motor Trade Association (MTA) voucher as compensation for their time.

2.2.2 **Demographics**

A total of 49 listeners with NH participated in the current research. However, on inspection of the data, a number of participants were highlighted to have experienced particular difficulty with the task, leading to bias in the data set. Due to such bias, the data collected from these participants was excluded from the final analyses, resulting in a final total of 42 participants. Participants were randomly assigned to one of the four listening conditions in the current research – closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise. Table 4 outlines the participant demographics for each condition.

Table 4.

Participant Demographics

	n	***M* Age (years)**	***M* PTA (dB) L**	**R**	**Gender**
CC	15	21.80	2.00	3.60	*n* M = 8 *n* F = 34
OC	10	23.20	1.36	2.27	
CB	9	25.50	1.50	2.70	
OB	8	28.00	4.40	3.13	
Total	42	24.63	2.31	3.02	

Note. CC = Closed set, Constant noise, OC = Open set, Constant noise, CB = Closed set, Babble Noise, OB = Open set, Babble Noise, n = number of participants, M = mean, PTA = pure tone average; R = right ear; L = left ear; M = Males, F = Females.

2.3 Stimuli

Sentence stimuli were presented at a constant level of 65 dB SPL amongst either the constant or babble masking noise. In order to ensure the accuracy of the level at which the signal was presented calibration was conducted, for which the procedure utilised is outlined in the following section. The lists were presented at two SNRs for each condition: - 13.9 dB SNR and -7.7 dB SNR for the constant noise condition and -14.3 and -7.6 dB SNR for the babble noise condition. These SNRs were selected in order to approach the pair of comprise, based on the literature suggesting this method to provide accurate and efficient estimates of SRT (Brand & Kollmeier, 2002; Ozimek et al., 2010). The SNRs were randomly assigned to half of the sentences in each test list for each condition to ensure that there were equal numbers of sentences presented at each SNR.

2.3.1 **Calibration of the Signal**

Initial assessments of the signal presentation level revealed inaccuracies that occurred from the method through which the intensity was measured. It became apparent that the measurements were affected by the silences between words and sentences, thus lowering the average signal level. This resulted in a subsequent increase in intensity due to the way in which the software attempted to compensate for this decrease in the overall presentation level. A precise measure of the signal level was required for the sentences designed for use in each noise type, following the removal of such gaps.

The recalibration procedure was conducted using a GRAS ISO 4869-3 Hearing Protector Test Fixture Type 45CA, fitted with a standardised artificial pinnae, with a 32-second averaging time. The signal was presented via a Brüel & Kjaer type 3560 C pre-amplifier and the differences were evaluated using version 17.1.1 of the Brüel &

Kjaer PULSE Labshop fast track software**.** Figure 7 illustrates the set up utilised for the recalibration procedure in the current research.

Figure 7. GRAS ISO 4869-3 Hearing Protector Test Fixture Type 45CA fitted with a standardised artificial pinnae.

The findings revealed the signal to be 3.9 dB SPL louder than the constant noise and 3.8 dB SPL louder than the babble noise. The 0.1 dB SPL difference between the two noise types was attributed to measurement error, given the identical procedure utilised to obtain the measurements for each noise type. Therefore, a 3.85 dB SPL difference between the signal presentation level and the two types of noise was identified. In order to account for this difference, 3.85 dB SPL was added to each SNR obtained in the current study for each noise condition, and retrospectively to McClelland's (2014) data described throughout this manuscript.

2.3.2 Generation of New Sentence Lists

As part of McClelland's (2014) work, 30 lists of 20 sentences were generated for both the constant and babble noise conditions. These lists were constructed manually (i.e. by trial and error) in Microsoft Excel. Because the words appeared the same number of times in each of the constant noise lists, the mean s_{word} was proposed to be identical for each list. In an attempt to preserve the reliability, and maximise the sensitivity, of the measure, two words were removed from the babble condition – "wins" and "shirts" (McClelland, 2014). The rationale behind the removal of these words was due to the abnormal psychometric function produced and the degree of adjustments required being deemed excessive, respectively (McClelland, 2014). Accordingly, the mean s_{word} varied slightly across the lists in the babble noise condition, however such differences were not considered to be of significance (McClelland, 2014). With regards to the sentence-specific slope ($s_{sentence}$), the mean varied across sentences – with some steeper and some shallower – for both noise conditions due to the words that appeared in each. The goal of achieving a high degree of similarity in the SD of the $s_{sentence}$ is to ensure that the lists are as similar as possible. The descriptive statistics for each of the lists appear in Tables 10 and 11 of McClelland (2014, pp. 95-96).

The sentence lists discussed above were generated using only the auditory psychometric properties as criteria however, for the UCAMST, consideration of the auditory-visual component was also necessary. Some of the sentences may have revealed suitable psychometric properties in the auditory-alone condition, but may contain poor quality visual transitions. As part of the current study, new sentence lists were generated to maximise the visual quality of the transitions between sentences, while maximising the SD of the $s_{sentence}$ between lists. Based on the calculated "pixel difference value" between successive video frames on either side of an edited transition, Trounson (2012) classed judder magnitudes into "tier groups" with tiers 0 and 1 classed as "no judder", and tiers 2 through to tier 6 having judders of increasing magnitude. Rather than generating 30 lists of 20 sentences, an alternative approach was taken whereby 16 lists of 10 sentences were generated such that they could be combined in pairs randomly during testing. The software was written to iteratively generate a set of sentence lists according to the following methodology. For the constant noise, the 10 sentences in each list contained one occurrence of each word. As mentioned, the words “wins” and “shirts” were removed from the babble noise condition, thus the verb and noun columns contained at least one occurrence of the other nine words, with one word, selected at random, appearing twice. The 10 words in each column were shuffled randomly to produce 10 sentences. A sentence was rejected if it contained one or more transitions with a judder magnitude of tier 3, or higher, or three transitions of that were classified as tier 2. Various methods were used to save calculation time – for example, if, say, the first six sentences of a list were accepted, but the seventh was rejected, then rather than abandoning the entire list, the remaining four sentences were re-shuffled to form new ones, and then these sentences were tested against the criteria described until they were deemed acceptable. Each of the sentence lists were added to a stack containing the lists, which was continually sorted by the SD of the s_{word} values (i.e. lists with the lowest SD were placed at the top of the stack). Due to their different psychometric properties, the lookup tables for the s_{word} values of the constant noise and babble noise were different. As each new list entered the stack it was compared to the sentences already present and was inserted at the appropriate place. If a list contained a sentence already in the stack it was rejected unless the SD of the incoming list was lower than the duplicate, in which case it was the duplicate list that was deleted from the stack. This process was repeated 100,000 times until a stack of 20 unique lists with considerably lower SDs was produced. The visual components of the sentences in these 20 lists were inspected for judder by two observers (i.e. the author and the primary supervisor). This process identified four lists that contained a large number of sentences with subjective judders. Accordingly these lists were deleted, leaving 16 lists of 10 sentences suitable for use in each condition.

2.4 Experimental Instrumentation

The initial hearing screening was conducted in a sound-treated audiologic test booth at the University of Canterbury Speech and Hearing Clinic (Christchurch, NZ). In order to obtain audiometric hearing thresholds, participants were presented with octave pure tones across the frequency range 250 – 8000 Hz via a calibrated Gradson-Sadler GSI clinical audiometer. Pure tones were presented via Telephonics TDH-50P supra-aural headphones worn by participants, who indicated hearing the tone by pressing a push button linked to the audiometer.

The experimental procedure was conducted in a research laboratory at the University of Canterbury Department of Communication Disorders (Christchurch, NZ). Participants were seated in the laboratory either alone or with the researcher, dependent upon the condition to which they were assigned (i.e. participants in the closed set condition were alone, given that this test format enables participants to self-administer the test by selecting the appropriate response on a touchscreen computer). The UCAMST software was developed via LabVIEW and was designed for use on a laptop computer. The current procedure utilised a Toshiba Tecra laptop, connected to an ēlo touch-sensitive monitor (ēlo ET1715L, Tyco Electronics, CA, USA) that was used by the participants or, in the open set condition, the researcher to select the appropriate responses. Sentence stimuli and masking noise were presented through Sennheiser HD280 Pro circumaural headphones (64 Ω impedance). The data collected were investigated through the generation of intelligibility functions using version 14.4.7 of Microsoft Excel. All of the statistical analyses performed on the data were conducted using the IBM Statistical Package for the Social Sciences (SPSS, version 21).

2.5 Scoring Procedure

The current study employed word scoring of the UCAMST sentences. Based on the findings of McClelland (2014), rather than scoring with regards to whether the fragment or sentence was correctly identified, this procedure calculated the number of words correctly recognised in each sentence. Therefore, each participant was awarded a score out of five for each sentence, which reflected the number of words that were correctly recalled.

2.6 Experimental Procedures

Prior to completing the procedure, each participant was asked a series of questions regarding their perception of their hearing ability and whether they had any history of health concerns directly related to their hearing, such as recurrent ear infections. Once this was completed, otoscopic examination was conducted to ensure all participants' external ears were free of excessive wax or debris that may have impacted the audiometric results obtained. As described, participants were seated in a sound-attenuated booth to complete the pure tone audiometry. Participants were instructed to only respond when they heard a tone, even if it was only slightly audible. The results obtained from this aspect of the study were explained to each participant before continuing to the experimental task. Any participants identified to have a HI were informed that they did not meet the inclusion criteria of the study and were given information regarding the follow-up procedures (refer to p. 1 of Appendix C.1 for details).

The tasks required of the participants differed depending on the experimental condition to which they were assigned. Participants in the closed set group were seated alone in a quiet room in front of a touch responsive computer monitor displaying the 50-word matrix from which the sentences were selected. The layout of the response panel that was made visible to participants after each trial is depicted in Figure 8.

Figure 8. Closed set response matrix displayed to participants following each trial.

Verbal instructions were given to each participant explaining that they would hear, via a set of headphones, a series of sentences in noise of varying volume. They were informed that their task was to identify the sentence heard by selecting the corresponding words from each column on the touch screen. All participants in this condition were informed to speculate what the sentence may have been when uncertain, in order to progress to the next trial.

The procedure for participants in the open set condition were comparable to those in the closed set condition, except that the presence of the researcher was necessary in order to score performance. Participants assigned to the open set task were also seated in a quiet room but were required to face away from the computer screen that displayed each sentence as it was presented (depicted in Figure 9). Similar verbal instructions were given to participants, but differed to those given in the closed set condition given that participants in the open set condition were required to identify the sentences heard by verbally responding. Participants in this group were also encouraged to guess when uncertain, and to respond with any words identified in the instance that the entire sentence was not recognised.

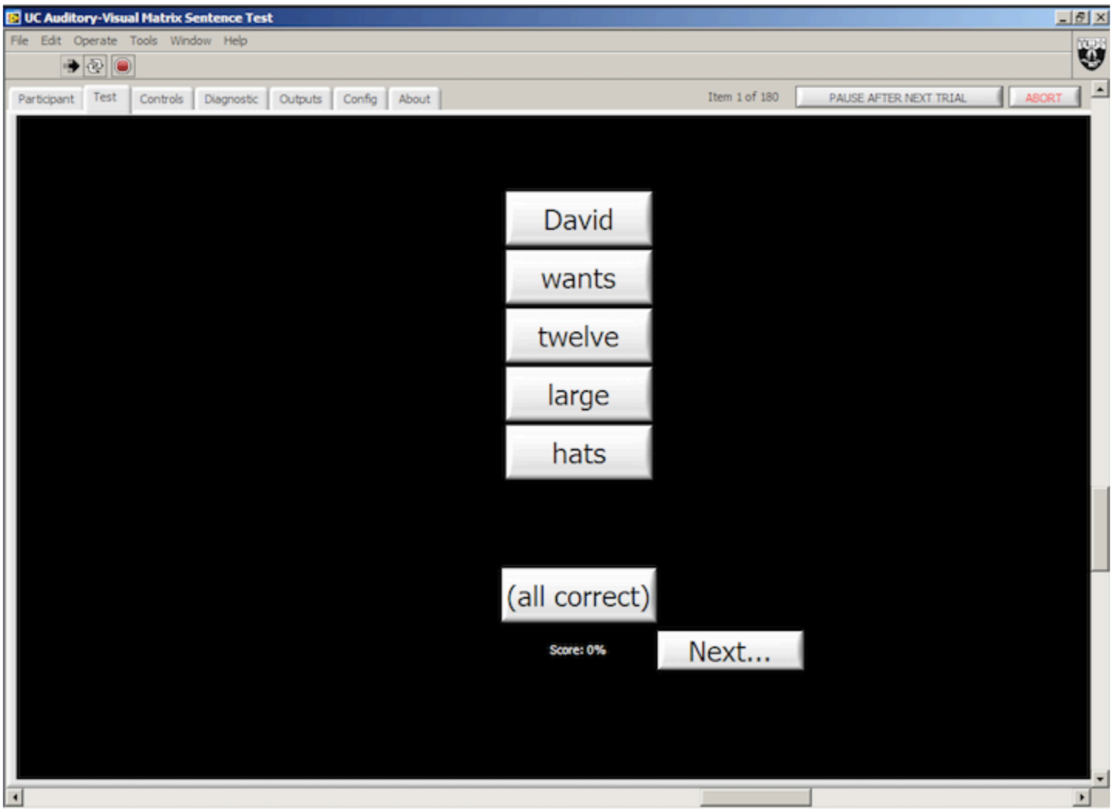

Figure 9. Open set response matrix used by the researcher to select words identified by participants

Irrespective of the condition, all participants were presented with 20 practice sentences (i.e. two lists) in order to ensure comprehension of the task and to allow familiarity with the test format and material. Following the practice sentences, all participants completed 160 test sentences (i.e. 16 lists) from which their performance was utilised as data in the current analyses. Due to the concentration required in completing the task, participants in both conditions were encouraged to take rest breaks as required. Excluding such breaks, the complete procedure took approximately 60 minutes for participants to complete, irrespective of condition.

2.7 Statistical Analyses

Prior to analysis the data were first examined for potential sources of bias that may violate the assumption of normality (i.e. significant skewness or kurtosis, or any outlying data), in order to determine whether parametric analyses could be performed. In each analysis, significant bias was revealed in the data and therefore non-parametric (i.e. assumption-free) analyses were implemented to test each of the study hypotheses. Specifically, to test hypotheses (1) to (4), a Kruskal-Wallis one-way ANOVA was used. A 5-minute time-out was used for the exact *p*-value calculations. If the time-out occurred, the asymptotic significance level was reported.

Following the hypothesis testing for hypotheses (1) to (4), post-hoc (i.e. observed) power (1-β) and estimated effect size (η^2) were calculated within a univariate ANOVA. All significant Kruskal-Wallis ANOVAs were followed by examining the pairwise comparisons within the univariate ANOVA.

2.7.1 **Planned Analyses**

A RM-ANOVA was planned to assess hypotheses (5a) and (5b). As previously stated, however, there was a lack of normality in the distribution for slope and SRT. Further, there were also significant outliers in the data for both variables. When attempting a RM-ANOVA, Box's Test was revealed to be significant ($p < .001$) for both slope and SRT, indicating that the covariances of the variables were significantly different. In addition, sphericity could not be assumed for either slope or SRT. While it would have been possible to use a Greenhouse-Geisser corrected F-ratio in the analyses, given that the data violated multiple assumptions underlying this analysis, the hypotheses were tested using non-parametric analyses. The power for both analyses (in the parametric environment) was > .999. The observed effect size for the slope analysis was $\eta^2 = .194$ and for the SRT analysis was $\eta^2 = .336$.

CHAPTER THREE:

RESULTS

3.1 Overview

This chapter presents the results of the analyses performed on the data collected in the current study. First, the results of the list equivalence assessment, described by research questions (1) and (2), are presented and compared between conditions. Next, the results to research question (3) are described. The implications of the findings are discussed in Chapter Four.

3.2 List Equivalence Results

The results of the analyses aimed at testing hypotheses (1) to (4) are presented in Table 5. Generally, the data supported the study hypotheses, however two instances were identified whereby the data either did not support the hypotheses, or the statistical power was not great enough to determine whether that was the case.

Table 5.

χ^2 and p-values for the Kruskal-Wallis one-way ANOVA on sentence lists in each of the four conditions

Condition	Variable	χ^2	p	1-β	η^2
Closed Constant	Slope	12.98	0.604	0.668	0.100
	SRT	8.82	0.887	0.321	0.051
Open Constant	Slope	19.46	0.194	0.847	0.552
	SRT	9.81	0.832	0.863	0.575
Closed Babble	Slope	31.74	0.007	0.818	0.128
	SRT	20.81	0.143	0.882	0.145
Open Babble	Slope	34.27	0.003	0.940	0.168
	SRT	13.38	0.573	0.771	0.119

Note. Degrees of freedom = 15.

Prior to performing the analyses, descriptive statistics were examined for each of the lists in each condition. These values are provided in Table 6.

Table 6.

Means and Standard Deviations of the slope and SRT of the lists designed for use in each condition of the UCAMST

	Condition															
	Closed Set, Constant Noise				Open Set, Constant Noise				Closed Set, Babble Noise				Open Set, Babble Noise			
	Slope (%/dB)		**SRT (dB SNR)**		**Slope (%/dB)**		**SRT (dB SNR)**		**Slope (%/dB)**		**SRT (dB SNR)**		**Slope (%/dB)**		**SRT (dB SNR)**	
List	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
1	11	3	-10.82	1.06	11	3	-9.17	0.58	6	2	-8.69	2.08	49	41	-8.46	1.09
2	11	4	-10.55	1.28	31	38	-8.59	1.35	8	2	-9.00	2.11	16	22	-6.85	2.26
3	11	4	-10.57	1.08	28	37	-8.56	0.85	9	5	-10.21	1.98	22	31	-7.20	0.88
4	11	3	-11.08	1.11	27	33	-9.08	1.11	6	6	-7.72	4.24	48	41	-8.00	1.15
5	12	3	-10.21	0.57	27	34	-8.13	1.24	7	1	-9.62	1.08	30	37	-4.34	6.86
6	12	3	-11.11	1.36	36	40	-8.96	1.35	10	2	-10.08	0.85	15	23	-7.49	1.88
7	10	3	-10.55	1.49	37	40	-8.82	1.20	7	1	-10.00	0.80	41	41	-7.61	0.72
8	14	3	-10.41	1.27	52	42	-8.20	0.89	16	24	-9.55	1.00	24	37	-6.02	3.89
9	11	2	-10.67	1.04	44	39	-8.88	1.39	10	4	-10.10	1.50	22	31	-6.53	2.81
10	20	26	-10.99	1.25	20	26	-8.83	1.40	6	3	-8.43	2.70	47	41	-7.51	1.26
11	11	3	-10.74	1.12	27	35	-8.66	1.42	7	3	-10.17	2.33	33	39	-7.75	0.94
12	13	4	-10.64	0.89	37	38	-8.83	1.19	10	5	-10.41	1.37	15	22	-5.59	5.52
13	12	2	-10.35	0.99	36	43	-8.47	0.92	6	4	-7.44	5.95	25	31	-7.25	1.11
14	12	2	-10.35	0.99	37	39	-8.72	1.15	7	4	-8.66	2.06	46	39	-7.69	0.79
15	11	4	-10.53	1.40	32	37	-8.44	1.15	8	3	-10.42	1.58	19	31	-5.48	5.49
16	11	2	-10.87	1.20	18	23	-6.51	7.75	8	3	-9.48	2.26	66	41	-7.54	0.59

Hypothesis (1) – There are no significant differences between the stimulus lists in the *closed set, constant noise* *condition for (a) slope and (b) SRT:*

As noted in section 2.7, the Kruskal-Wallis ANOVA was utilised in order to determine whether the lists in a given condition were equivalent to one another. As outlined in Table 5, the data supported this hypothesis for both slope and SRT, therefore indicating that no significant differences were found between the stimulus lists designed for use in the closed set, constant noise condition with regards to either the slope or SRT. This finding is illustrated in Figure 10, where the intelligibility functions for each list are shown and the lack of variation between stimulus lists can be recognised.

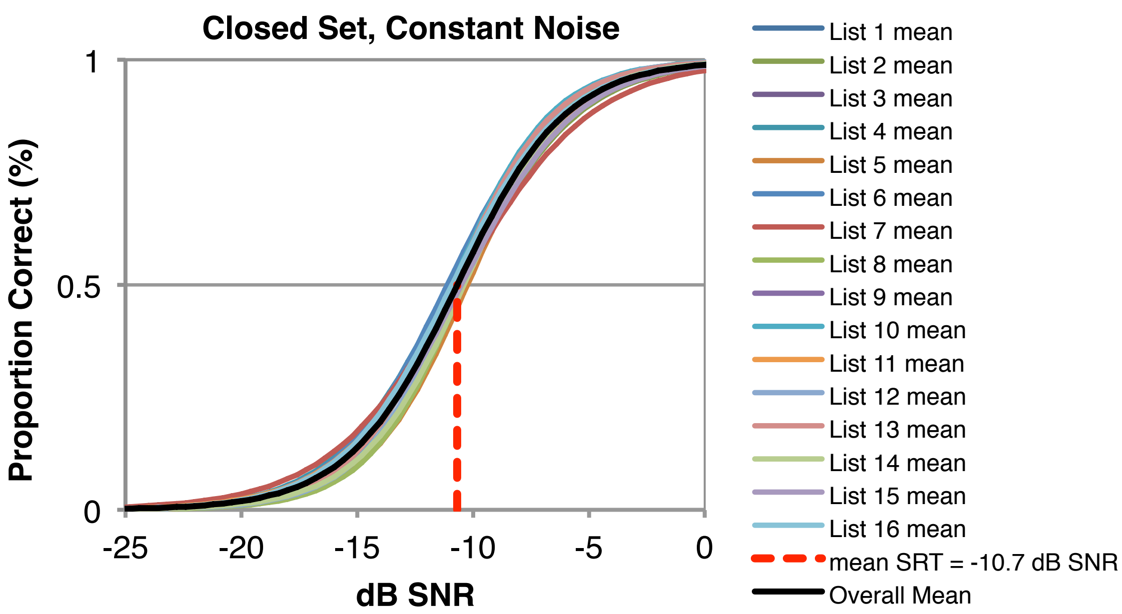

Figure 10. Intelligibility functions of the lists designed for use in the closed set, constant noise condition

Despite this encouraging finding, the post-hoc power analysis revealed that there was insufficient power in this analysis to identify a difference between the lists, if such a difference exists (defined as $1-\beta > .80$). Based on this, the current study cannot determine whether the statistical analysis was successful in capturing the variance between the lists designed for the closed set, constant noise condition of the UCAMST. Therefore, while it appeared that the data were in support of hypothesis (1), it cannot be ascertained whether this finding resulted from the lack of power preventing any present differences from being detected in the analyses.

Hypothesis (2) – There are no significant differences between the stimulus lists in the open set, constant noise condition for (a) slope and (b) SRT:

The Kruskal-Wallis one-way ANOVAs conducted on this data revealed no significant differences between the stimulus lists designed for use in the open set, constant noise condition with regards to slope and SRT (refer to Table 5). The similarities described by these analyses are depicted in Figure 11 whereby the overlap between the list-specific intelligibility functions can be realised.

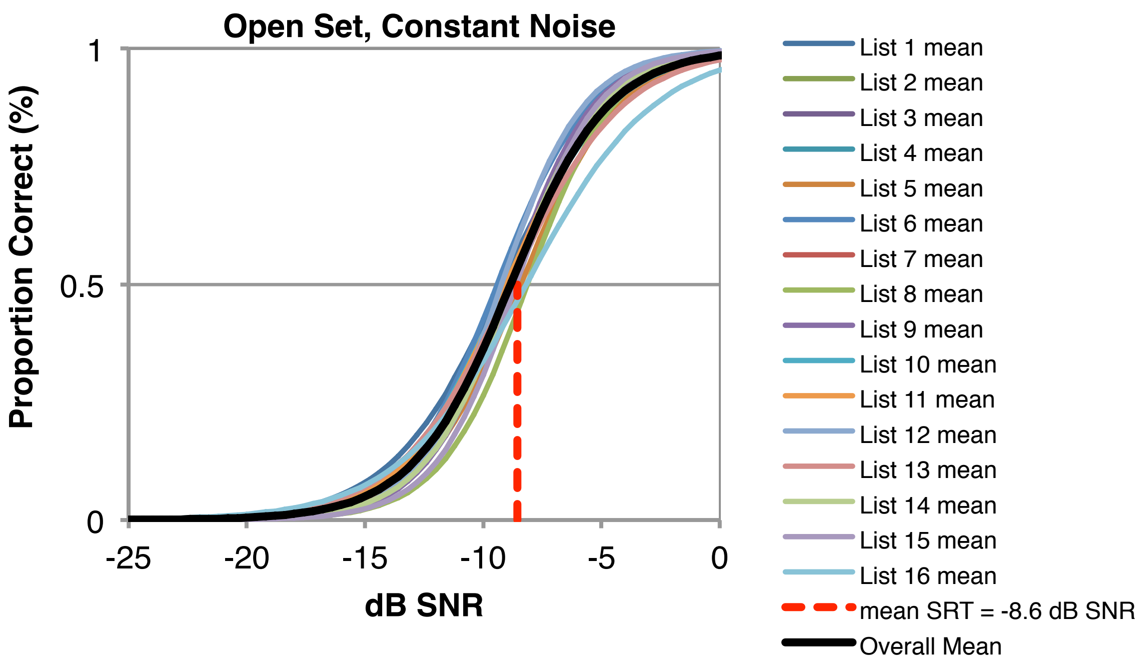

Figure 11. Intelligibility functions of the lists designed for use in the open set, constant noise condition.

In addition to this finding, the post-hoc power analysis conducted determined that both analyses had sufficient power (defined as 1-β > .80) to detect an existing difference, therefore indicating the variables to be effective in describing the variance between the lists designed for use in the open set, constant noise condition of the UCAMST. From this it can be concluded with confidence that the data supported hypothesis (2) of the current study, suggesting that the lists designed for this condition are equivalent to one another.

Hypothesis (3) – There are no significant differences between the stimulus lists in the closed set, babble noise condition for (a) slope and (b) SRT:

As described in Table 5, analysis of the data from the closed set, babble noise condition revealed, with sufficient power (defined as 1-β > .80), support of hypothesis (3) with regards to the SRT across lists. This finding indicates that there were no significant differences between lists designed for use in this condition in terms of the SNR at which SRT can be estimated. The data did not however, support this hypothesis with regards to the slope of the list functions, indicated in Table 5 by the significant result obtained. Post-hoc pairwise comparisons are shown in Table 7.

Table 7.

p-values for the pairwise comparisons of the slopes of lists in the closed set, babble noise condition

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1		.38	.29	.94	.74	.19	.68	**.001**	.18	.78	.62	.13	.92	.77	.41	.39
2			.72	.42	.59	.66	.64	**.01**	.65	.55	.70	.51	.44	.56	.96	.99
3				.24	.36	.94	.41	**.03**	.92	.34	.46	.77	.25	.34	.67	.71
4					.79	.21	.73	**.001**	.21	.83	.76	.14	.97	.83	.45	.42
5						.33	.93	**.003**	.32	.95	.87	.23	.81	.96	.62	.59
6							.34	**.04**	.98	.29	.41	.83	.22	.30	.62	.65
7								**.003**	.35	.89	.93	.26	.75	.90	.68	.65
8									**.04**	**.002**	**.004**	.06	**.001**	**.002**	**.01**	**.01**
9										.29	.40	.85	.22	.23	.61	.64
10											.82	.21	.86	.99	.58	.55
11												.30	.69	.83	.74	.71
12													.15	.22	.48	.51
13														.85	.47	.44
14															.99	.56
15																.96

Note. ***Bold*** *indicates a significant difference between the slope of the test lists.*

As depicted in Table 7, the slope of list 8 differed significantly from the slope of every other list, except list 12, $p = .06$, $\eta^2 = .128$. No other significant differences between the slopes of the lists were revealed, suggesting therefore that the presence of list 8 in the analysis may have influenced the significant finding obtained.

Together these findings can be visualised in figure 12 whereby the similarities between the mean SNR for each list is illustrated alongside the variations between the slopes of the intelligibility functions described.

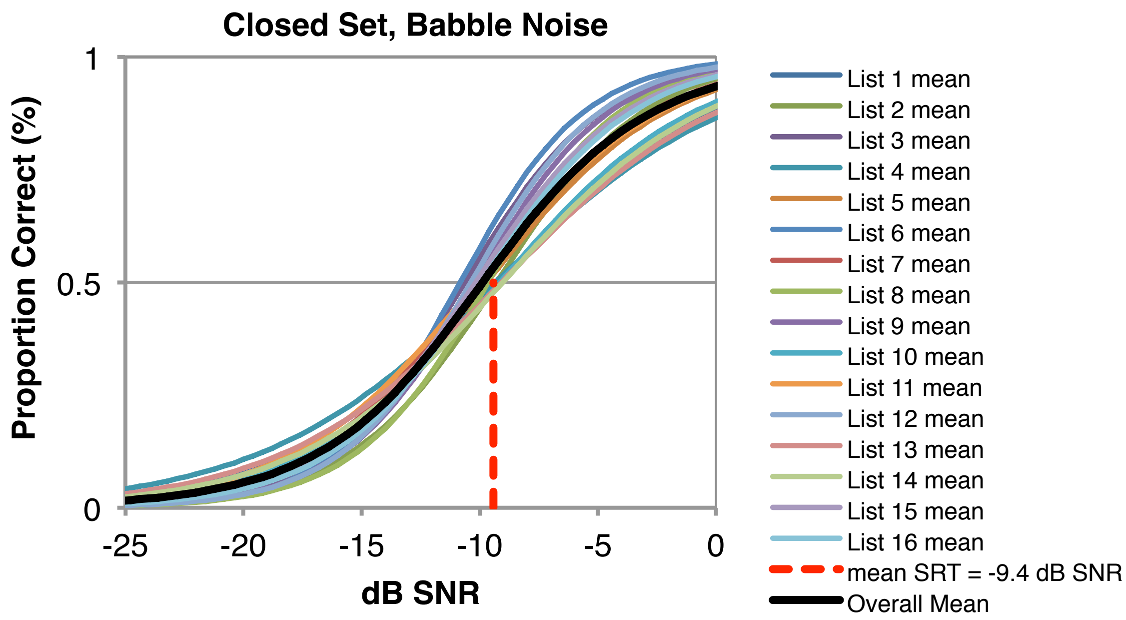

Figure 12. Intelligibility functions of the lists designed for use in the closed set, babble noise condition

Hypothesis (4) – There are no significant differences between the stimulus lists in the <u>open set, babble noise</u> condition for (a) slope and (b) SRT:

The Kruskal-Wallis one-way ANOVA conducted on the data revealed no significant difference between the stimulus lists designed for use in the open set, babble noise condition of the UCAMST with regards to SRT. Post-hoc power analyses revealed this analysis to have sufficient power (defined as 1-β > .80) to detect a difference between the lists, if such a difference exists. This finding therefore indicates that the variables were successful at summarising the majority of the variance between the lists in this condition. With regards to the slope of the lists, however, a significant difference was revealed, as noted in Table 5. The post-hoc pairwise comparisons conducted are shown in Table 8.

Table 8.

p-values for the pairwise comparisons of the slopes of lists in the open set, babble noise condition

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1		**.03**	.09	.95	.22	**.03**	.62	.11	.08	.93	.31	**.03**	.12	.88	.06	.26
2			.67	**.04**	.37	.97	.11	.59	.68	**.04**	.27	.97	.55	.06	.80	**.001**
3				.11	.64	.65	.23	.91	.98	.11	.49	.64	.87	.12	.85	**.005**
4					.24	**.03**	.67	.13	.10	.98	.35	**.03**	.14	.93	.07	.23
5						.36	.46	.72	.62	.25	.82	.36	.76	.28	.52	**.01**
6							.10	.57	.66	**.04**	.25	.99	.53	**.04**	.78	**.001**
7								.27	.22	.68	.61	.10	.30	.73	.17	.11
8									.89	.14	.56	.57	.95	.15	.67	**.007**
9										.11	.47	.66	.85	.12	.87	**.005**
10											.36	**.04**	.15	.95	.07	.22
11												.25	.60	.39	.38	**.03**
12													.54	**.04**	.78	**.001**
13														.17	.72	**.009**
14															.86	.20
15																**.003**

Note. ***Bold*** *indicates a significant difference between the slopes of the test lists*

These comparisons revealed that the slope of each of the lists, except list 7, differed significantly from the slope of at least one other list. Figure 13 presents these differences in addition to the equivalence of the SRT found across test lists designed for use in this condition. In combination, it can therefore be concluded that hypothesis (4) was only supported with regards to the SRT of the list functions.

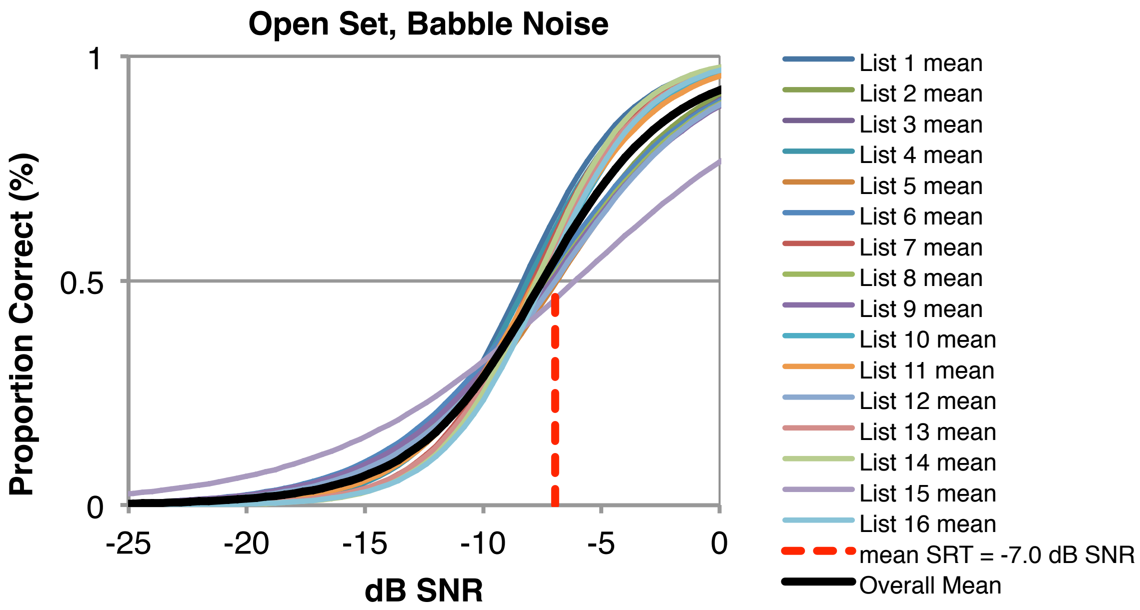

Figure 13. Intelligibility functions of the lists designed for use in the open set, babble noise condition

3.3 Condition Equivalence Results

Hypothesis (5a) – There is no significant difference between the _slope_ *for the four test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise):*

Friedman's related-measures two-way ANOVA was performed in order to analyse hypotheses (5a) and (5b). The results of this analysis indicated the slopes of the stimulus lists in each condition to be significantly different to one another, χ^2 (3) = 111.33, $p < .001$, thus not supporting hypothesis (5a). Wilcoxon signed rank tests were used to follow-up this finding. As outlined in Table 9, it appeared that significant differences ($p < .05$) existed for each comparison, except between the closed set, constant noise and the open set, babble noise conditions. Overall, with regards to slope, it cannot therefore be concluded that the lists designed for each condition are similar to one another.

Table 9.

Z-values of Wilcoxon signed rank test for the slope across the test conditions

	Closed set, Babble noise	Closed set, Constant noise	Open set, Babble noise	Open set, Constant noise
Closed, Babble		-9.60	-6.08	-9.61
Closed, Constant			-2.08 (.037)	-5.23
Open, Babble				-10.69

Note. All tests were significant at p < .001, except where noted in parentheses.

Hypothesis (5b) – There is no significant difference between the SRT (dB SNR) for the four test conditions (i.e. closed set, constant noise; open set, constant noise; closed set, babble noise; open set, babble noise):

The analyses conducted revealed significant differences between the SRT of the stimulus lists in each condition, $\chi^2 = 282.15$, $p < .001$. Therefore it can be concluded that the data did not support hypothesis (5b). Wilcoxon signed rank tests were implemented to conduct follow-up pairwise comparisons. The results, shown in Table 10, indicated that significant differences existed for each comparison ($p < .001$), therefore suggesting that the lists designed for use in each condition were not found to be equivalent to one another, with regards to SRT.

Table 10.

Z-values of Wilcoxon signed rank tests for the SRT across test conditions

	Closed set, Babble noise	Closed set, Constant noise	Open set, Babble noise	Open set, Constant noise
Closed, Babble		-6.33	-8.52	-10.95
Closed, Constant			-10.95	-10.82
Open, Babble				-8.43

Note. All tests were significant at p < .001.

The findings related to hypotheses (5a) and (5b) are presented in Figure 14, whereby the differences between the slope and SRT of the intelligibility functions of each of the conditions revealed are illustrated.

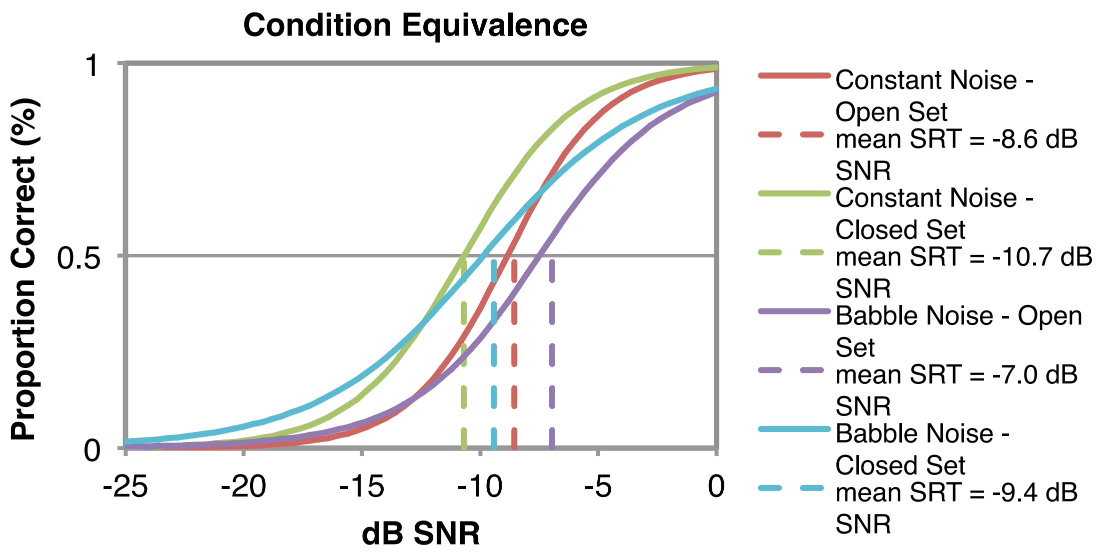

Figure 14. Intelligibility functions of each of the conditions of the UCAMST

3.4 Comparison of Results with Published MSTs

Hypothesis (6a) – There is no significant difference between the stimulus lists designed for use in the UCAMST and those of previously published MSTs (Dietz et al., 2014; Houben et al., 2014; Jansen et al., 2012; Øygarden, 2009; Ozimek et al., 2010; Puglisi et al., 2014; Wagener et al., 2003; Warzybok et al., 2015) with respect to (a) slope and (b) SRT.

In order to determine whether the UCAMST stimulus lists were equivalent to those designed for previously published MSTs, single samples *t*-tests were conducted. The results of these analyses revealed statistically significant differences between the UCAMST stimulus lists and those of the published measures included in the analyses (Dietz et al., 2014; Houben et al., 2014; Jansen et al., 2012; Øygarden, 2009; Ozimek et al, 2010; Puglisi et al., 2014;

Wagener et al., 2003; Warzybok et al., 2015) at the $p < .001$ significance level, with regards to both slope and SRT. One exception to this finding was that the mean SRT of the open set, constant noise condition was found to be similar to the mean SRT of the Danish MST (Wagener et al., 2003), $p = .384$.

Table 11 provides a comparison across the mean SRTs (dB SNR) and slopes (%/dB) of international MSTs, highlighting the differences described above.

Table 11.

Mean SRT and slopes of International MSTs

	MST	***M* SRT**	***M* Slope**	**Authors**
Closed Set, Constant Noise	Dutch	-8.4 ± 0.2	10.2 ± 0.9	Houben et al. (2014)
	French	-6.0 ± 0.1	14.0 ± 1.6	Jansen et al. (2012)
	Italian	-7.3 ± 0.2	13.3 ± 1.2	Puglisi et al. (2014)
	UCAMST	**-10.7 ± 0.2**	**10.6 ± 0.9**	**Current**
Open Set, Constant Noise	Danish	-8.4 ± 0.16	12.6 ± 0.8	Wagener et al. (2003)
	Norwegian	-6.0 ± 0.8	14.0 ± 1.6	Øygarden (2009)
	Finnish	-10.1 ± 0.1	16.7 ± 1.2	Dietz et al. (2014)
	Russian	-9.5 ± 0.2	13.8 ± 1.6	Warzybok et al. (2015)
	UCAMST	**-8.8 ± 0.4**	**12.3 ± 1.5**	**Current**
Closed Set, Babble Noise	**UCAMST**	**-9.8 ± 0.6**	**7.3 ± 1.4**	**Current**
Open set, Babble Noise	Polish	-9.6 ± 0.2	17.7 ± 1.6	Ozimek et al. (2010)
	UCAMST	**-7.4 ± 0.5**	**9.1 ± 2.1**	**Current**

Note. M = mean; ± (*x)* = SD. *The differences across the reported M SRT or M slope values of the UCAMST conditions were concluded to be attributable to rounding error.*

As shown, while a range of acceptable SRTs and slopes are evident across international versions, the degree to which the UCAMST differs from these versions can be realised. Additionally, the similarities between the SRT of the UCAMST and the Danish MST (Wagener et al., 2003) revealed in the analyses are depicted.

When considering the slope and SRT of international MSTs, Figure 15 enables the differences between the UCAMST and previously published versions to be visualised. For the most part, examination of Figure 15 reveals the UCAMST to have a shallower slope as compared to international versions, which may have implications regarding the accuracy of SRT estimations obtained when employing this measure. This, as well as other inferences made from the results, will be discussed in the following chapter.

3.5 Summary

To summarise, this section provides a review of the main results obtained in the current study:

(1) The lists in the constant noise condition were revealed to be equivalent with regards to a) slope and b) SRT. In contrast, significant differences were found between the lists designed for use in the babble noise condition with regards to slope, however were found to be comparable with regards to SRT.

(2) Accordingly, analysis across the four conditions of the UCAMST indicated significant differences to exist between the expected estimates of a) slope and b) SRT obtained in each condition.

(3) When compared to international versions, the UCAMST was revealed to differ from each of MSTs included in the analysis with regards to a) slope and b) SRT. Interestingly, one exception to this finding occurred between the UCAMST and the Danish MST (Wagener et al., 2003) whereby equivalence was observed with regards to SRT.

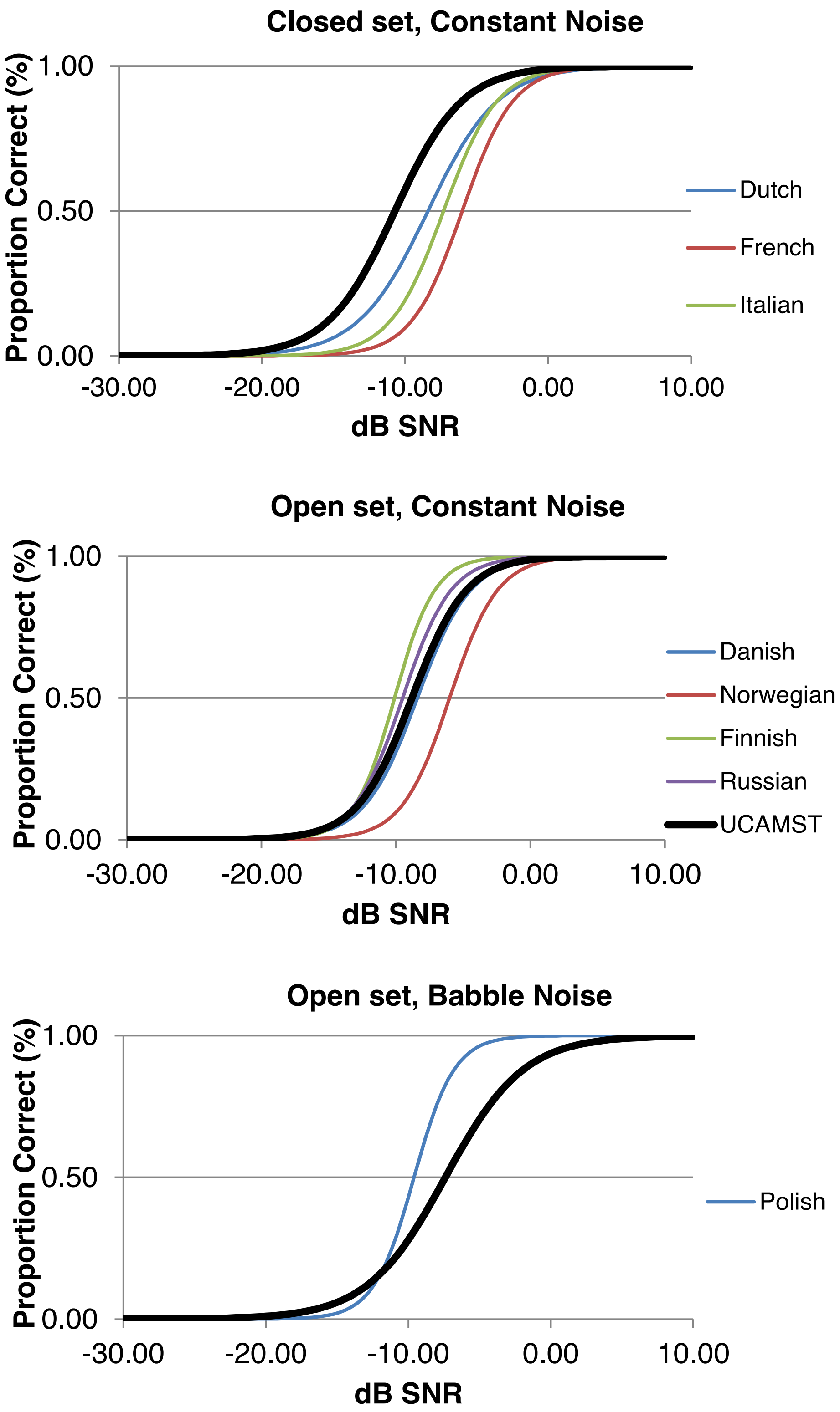

Figure 15. Comparison of slope across international MSTs.
Note. The intelligibility function of the Closed Set, Babble Noise condition is not shown, as it cannot be compared against an international MST due to the absence of this condition in international versions.

CHAPTER FOUR:

DISCUSSION

4.1 Introduction

The purpose of this research was to evaluate the difficulty of the test lists designed for use in the UCAMST in order to determine whether the lists were equivalent to one another. Subsequently, it was also of interest to examine whether the UCAMST stimuli were equivalent to previously published MSTs. The results of the list analyses revealed that while the lists designed for use in some conditions were equivalent, others were not. Further, the results of the condition analyses confirmed differences in performance based on the masking noise and the mode of presentation employed. Finally, comparison with internationally published MSTs revealed statistically significant differences between the UCAMST and such measures regarding both the SRT and the slope. In combination, such findings warrant further examination of the stimulus lists designed for the UCAMST and highlight the importance of the evaluation stage in developing new clinical and research tools. This chapter will discuss the findings with reference to the literature, outline the limitations of the study and consider the implications of such drawbacks, and suggest future areas of research.

4.2 Equivalence Measures

4.2.1 **List Equivalence**

The first cluster of hypotheses proposed to address this research question predicted that the lists designed for each condition of the UCAMST would be equivalent to one another with regards to the SRT and slope of the functions. The results of the analyses supported some, but not all, of the hypotheses. Evaluation of the constant noise lists produced non-significant results, therefore suggesting the lists designed for use amongst this type of noise, irrespective of presentation mode, to be equivalent to one another with regards to both slope and SRT. This finding is in accord with what was expected, based on the standard methodology employed across the development of MSTs and indicates the ability to use the list stimuli in the open and closed set modes interchangeably (Akeroyd et al., 2015).

Contrary to these hypotheses, evaluation of the babble noise lists revealed equivalence between the lists with regards to the SNR at which the SRT can be estimated from, but not with regards to the slope of the intelligibility functions. This result was found both in the open and in the closed presentation modes. While the slope of each of the lists in the open set babble noise condition differed from at least one other, the closed set condition revealed one list in particular to differ from the others – list 8. Based on the strength of the significant findings for this list, deletion of list 8 may improve the equivalence between the lists in this condition, as compared to altering the list. However, as they stand, the current findings suggest that when assessing speech recognition using the babble noise condition, the estimated SRT is likely to be comparable across list stimuli but that the slope of the intelligibility function may fluctuate based on the lists employed.

In combination, these findings have important implications with regards to the use of the UCAMST in both research and clinical settings. Equivalence across the slopes of the lists enables the administrator (i.e. the researcher or clinician) confidence in concluding a listener's SRT, despite the lists employed in the testing procedure. Therefore, while it is likely, based on the current findings, that a reliable estimate of a listener's SRT will be obtained when administering lists designed for use with constant noise, the same cannot be concluded for the babble noise lists.

The unexpected similarities between the babble noise lists may be ascribed to a number of factors relating to the methodology. First, an unforseen malfunction in the software occurred, affecting the selection of list stimuli for the babble noise conditions. Consequently, listeners in the babble noise conditions were presented with the constant noise stimulus lists in place of those designed for use amongst this type of noise. This may have had a considerable impact on the listeners' ability to complete the task, and consequently the resulting estimates of SRT and slope obtained for the current analyses, since the list stimuli were designed for exclusive use amongst each of the two noise types and were optimized in order to achieve a high level of homogeneity based on this premise. Thus, presenting list stimuli that were not optimized for use amongst the babble noise may have significantly influenced the findings of the current research. Second, the training effect associated with the UCAMST is yet to be determined. For the practice phase of the current study, the number of lists needed to stabilise performance was estimated based on the consensus provided in previous research (Dietz et al., 2014; Hochmuth et al., 2012). Therefore it may be possible that such training was inadequate in stabilising performance on the UCAMST, thus influencing the findings obtained. While this is improbable, based on the comparable procedures employed across studies, the importance of the training phase cannot be overlooked. Last, the small sample of participants that completed the babble noise conditions, as compared to the constant noise conditions, may have also contributed to the differences in homogeneity found between the lists designed for each condition. Further consideration will be given to the impact of such limitations in section 4.5 alongside suggestions for reducing such effects in future research.

4.2.2 **Condition Equivalence**

While estimates of the list equivalence offer useful information when evaluating a new measure of speech recognition, it is also of interest to determine the equivalence of the conditions. That is, whether the conditions provide reliable estimates of SRT and with comparable accuracy. The second cluster of hypotheses (i.e. 5a & 5b) were proposed for the current research in order to address this area of evaluation for the UCAMST.

Evaluation of the conditions did not support hypotheses (5a) and (5b) due to finding significant differences between the conditions with regards to both slope and SRT. This result therefore suggests that when estimating a given listener's SRT, both the slope of the intelligibility function and the SRT are likely to differ depending on the administration of the lists (i.e. the masking noise selected and the mode of presentation). One exception to this finding was noted between the closed set, constant noise condition and the open set, babble noise condition for the slope variable whereby no significant differences were found. Thus the data seemingly supports interchangeable use of these conditions, however due to the natural variations between performance in the open and closed set, this may not be advisable in practice.

Despite the mostly significant findings obtained between conditions, examination of the data presented a trend that may be expected, based on some of the literature, regarding the effect of the presentation mode on performance. Based on the current data, it appears that participants found the task more difficult when the stimulus sentences were presented in the open set mode. Similarly, Hochmuth et al. (2012) noted a significant difference between listener performance due to the mode through which the task was completed, revealing closed set testing to result in a higher SRT. While few studies have directly explored the differences between presentation modes for MSTs, during evaluation of the Polish MST, Ozimek et al. (2010) found opposing results to Hochmuth et al. (2012), whereby performance did not differ across the presentation modes. As previously discussed, this finding may have been attributable to the extensive one hour training procedure employed by Ozimek et al. (2010) to stabilise performance. This adaptation in the training regime may therefore explain the agreement between the findings of the current research and those of Hochmuth et al. (2012)

for which such training was not completed. In addition to these findings, a concurrent project (described in detail in section 4.6.3) investigating the use of the auditory-visual component of the UCAMST with listeners with HI, revealed that listeners had more difficulty on the task when the stimuli were presented in the closed set mode (André, in progress). It was suggested that this finding may have resulted from the greater cognitive demands of the task in this condition, as compared to in the open set condition, since the sentence needs to be retained throughout the time it takes to find the corresponding buttons in the base matrix (André, in progress). Therefore, it appears that ambiguity surrounding the effect of the presentation mode exists in the literature, suggesting that tailoring the use of the UCAMST conditions to the listener's capability may be required. This issue is considered further in section 4.6.4.

A further trend noted from inspection of the data was that participants appeared to experience more difficulty with the task when the stimulus lists were presented amongst the babble masking noise, SRT = -7 dB SNR and -9.4 dB SNR for open and closed set, respectively, as compared to constant noise, SRT = -8.6 dB SNR and -10.7 dB SNR for open and closed set, respectively. This finding is unique to the current study. Previous investigations of the differences in speech recognition due to the masking noise presented have generally reported babble noise to enable listeners, particularly those with NH, to take advantage of temporal and spectral dips (Peters et al., 1998; Wilson, Carnell & Cleghorn, 2007b). As described, these dips are thought to provide brief 'glimpses' of the target stimulus, thus improving SRT (Peters et al., 1998). Research investigating this phenomenon has supported this explanation for such discrepancies through varying the number of talkers included in the recording of the masker, in order to determine the optimum masker for estimating SRT. Simpson and Cooke (2005) examined the influence that gradually increasing the number of talkers from 1 to 512 had on speech recognition performance. Based on the findings, the researchers determined that as the number of talkers increased, the cues provided by dips became progressively less informative to listeners (Simpson & Cooke, 2005). Elsewhere, Van Engen and Chandrasekaran (2012) also postulated that performance generally declines as talkers are added to the masker, but that performance in six- to eight-talker babble is likely to be significantly better than when speech-shaped noise is utilised. Therefore, based on the literature, the results of the current study revealing performance to be better in the constant noise condition, as compared to the babble noise condition, is an unexpected finding.

One possible explanation for finding this conflicting trend may be due to the use of non-optimised test lists in the babble noise condition. As described, the lists were designed for use exclusively with the noise type for which they were optimised. Therefore, it is possible that presenting the lists designed for use in the constant noise condition with the babble masker may have had a significant influence on the participants' performance in this condition, and consequently the results of the current research.

While it is conceivable that finding enhanced performance in the presence of the constant noise masker, as compared to the babble noise, may have resulted from the use of non-optimised lists in the babble noise condition, it is also plausible that the process through which the masker interferes with the signal may have contributed to this finding. When speech is embedded in background noise, there are two main ways that noise can mask the speech signal – through either energetic or informational masking (Arbogast, 2003; Lidestam, Holgersson & Moradi, 2014). Energetic masking is thought to be produced by non-speech sounds whereby some portion of the masker energy falls within the same auditory filter as the signal energy (Myerson et al., 2016). In addition to energetic masking, informational masking is thought to occur in the absence of, or in addition to, spectral overlap between the signal and the masker, leading to high levels of uncertainty regarding the target stimulus or masker (Arbogast, 2003; Myerson et al., 2016). It has been suggested that the differences between these types of maskers reflect the portion of the auditory system at which the interference occurs (Myerson et al., 2016). It is thought that energetic masking interferes with processing at the peripheral level (i.e. in systems up to, and including, the auditory nerve) and that informational masking effects processing at higher levels (i.e. processes in the central auditory system) (Francart et al., 2011; Myerson et al., 2016). In accordance with this view, some researchers refer each type of masking as peripheral and central masking, respectively (Myerson et al., 2016; Wilson, Trivette, Williams & Watts, 2012).

Much research has demonstrated the challenge associated with determining whether the detrimental effect of noise on speech recognition can be attributed to energetic or informational masking (Lidestam et al., 2014; Wilson et al., 2012). Still, research efforts from Sperry, Wiley and Chial (1997) demonstrated separation of such masking effects through comparison across three types of competing background noise. The results led to the conclusion that meaningful and non-meaningful speech competitors are likely to give rise to greater degradation in performance than a non-speech competitor consisting of only the spectrum (Sperry et al., 1997). Similarly, more recent research implemented the HINT sentences (Nilsson et al., 1994) in order to simulate everyday listening environments and provided evidence that the number of talkers present in the background noise may influence the occurrence of informational masking (Hornsby, Ricketts and Johnson, 2006). Results showed that when the number of talkers was relatively small (i.e. two) informational masking effects may have been obscured by energetic masking (Hornsby et al., 2006). Conversely, as the number of talkers increased to seven, the effects of both informational and energetic masking were observed (Hornsby et al., 2006).

In view of such findings, consideration of the impact of informational masking on the present findings is justified. It is possible that the poorer than expected performance on the task in the babble noise condition, as compared to the constant noise condition, may be explained by the occurrence of informational masking. Accordingly, the difficulty with the task in this condition may be due to the greater levels of uncertainty associated with this type of masking (Myerson et al., 2016).

Despite this unforseen finding, one expected finding in the current study related to the sensitivity of the measure. Overall, the slope of the intelligibility functions of the babble noise conditions were found to be shallower – 9.1 %/dB and 7.3 %/dB for open and closed set, respectively – than those of the constant noise conditions – 12.3 %/dB and 10.6 %/dB for open and closed set, respectively. This finding is consistent with the differences noted in the literature between the slopes of each noise condition (Francart, 2011; Wagener & Brand, 2005). This is an important observation as this difference between the two types of noise is thought to have implications regarding the application of each test condition. A steeper slope signifies that a small change in SNR would yield a large change in SRT, thus denoting a highly sensitive measure (Theunissen, Swanepoel & Hanekom, 2009). Use of highly sensitive measures is thought to be advantageous in clinical settings where the schedule of clients often poses time constraints, under which administration of a large battery of tests is required. It has been noted that highly sensitive measures provide an accurate yet efficient method of estimating a listener's SRT, thus making such measures suitable in this environment (Ozimek et al., 2010). Alternatively, and as described, it is thought that babble noise may have higher face validity than constant noise, due to the more accurate representation of everyday listening contexts whereby multiple speakers are often present (Wilson et al., 2007a). Therefore, finding differences between the two noise types similar to those documented in the literature is an encouraging result of the current study as the application of the conditions may be guided by previous suggestions. Based on such work it may be advisable that selection of a test condition from the UCAMST is directed by the objective of administering the test.

In combination, the findings of the condition evaluation phase suggest that the UCAMST conditions should not, at this stage, be used interchangeably as it is unlikely that the results obtained would be comparable across conditions. It is possible, however, that the findings may have been subject to the limiting effects of the factors described in the previous section, for which the implications will be discussed in a subsequent section.

4.3 Comparison Across International MSTs

In order to infer whether the UCAMST was in accordance to previously published MSTs, it was of interest to the current research to evaluate across international versions. The analyses conducted revealed differences to occur between the UCAMST and each of the international versions to which it was compared. As previously acknowledged, one exception to this result was the equivalence revealed between the UCAMST and the Danish MST (Wagener et al., 2003), with regards to SRT. Overall, the findings of the current study therefore assert that speech recognition results obtained from the UCAMST are not yet comparable to those gathered via international MSTs.

Previous comparisons across international versions of the MST have revealed acceptable differences between the reference SRTs of each version (Kollmeier et al., 2015). The French (Jansen et al., 2012) and Norwegian (Øygarden, 2009) versions are reported to have the highest SRTs of -6 dB SNR, whereas the Finnish MST (Dietz et al., 2014) is thought to have the lowest SRT of -10.1 dB SNR. Therefore, a spread across the reference SRTs of international versions of 4.1 dB SNR is apparent (Kollmeier et al., 2015). In accordance with this observation, it is noteworthy that while the UCAMST stimulus lists for each condition were revealed to be statistically different from previous MSTs, examination of the intelligibility functions associated with each reveals such differences to be marginal, particularly for the open set constant noise condition.

Each of the published MSTs were developed using prescribed methodology and therefore various explanations have been postulated to account for such variation. First, unique language attributes, such as the frequency at which phonemes are produced, have been hypothesised to have an influence on the reference SRT (Kollmeier et al., 2015). For example, it is thought that for Slavic languages, such as Russian and Polish, that contain more high frequency phonemes, masking may be more difficult, possibly leading to lower SRTs (Kollmeier et al., 2015). Second, speaker characteristics, such as gender, have also been proposed as a possible contributor to the variation in the SRTs of international MSTs. Wagener et al. (2014) investigated this phenomenon using the versions of the German MST that were developed using a male speaker (Wagener, Brand & Kollmeier, 1999) and a female speaker (Wagener et al., 2014). The findings detected a difference in reference SRT of 2.2 dB SNR, thus supporting the potential influence that speaker characteristics may have on the homogeneity across versions (Wagener et al., 2014). Such evidence may be relevant to the current research findings. For example, both the Norwegian (Øygarden, 2009) and the Polish (Ozimek et al., 2010) MSTs employed male speakers, compared to the UCAMST whereby a female speaker was employed. Thus, based on the evidence provided by Wagener et al.'s (2014) work, it is possible that a proportion of the difference between the UCAMST and previous MSTs may be ascribed to this difference alongside the aforementioned limitations of the current research.

Finding international MSTs to be as homogenous as possible is of critical importance to the standardisation of therapy indications. That is, ensuring the consistency of result interpretation and hence the subsequent treatment options provided to clients across research centres and clinics (Bilger et al., 1984). Consistency in the interpretation of results is of particular importance to languages that are frequently spoken worldwide, such as Russian, French and Spanish, as the use of these tools is likely to be extensive (Dietz et al., 2014).

Further, advancements towards the ability for clients to move across the European Union (EU) for the purpose of accessing various health care schemes are currently occurring (Palm & Glinos, 2010). This shift aims to enable clients to have the ability to access healthcare outside of their home state when traveling or in order to receive superior quality, or more affordable, healthcare services (Palm & Glinos, 2010). Such developments also justify the importance of the standardisation of tests in order to assure the accurate interpretation of results by healthcare providers. Despite the unique qualities of NZ English restricting the use of the UCAMST to people in NZ, the significance of standardisation across MSTs remains as consistency across clinical and research settings is vital. Therefore, the results obtained in the current research require consideration in future research in order to achieve a measure that is comparable to international versions.

4.4 Summary

Overall, the results of this research necessitate further investigation of the UCAMST stimulus lists in order to defend its use. While encouraging results were uncovered, two major limitations arose in the general findings – the inability to compare estimates of speech recognition between lists and conditions of the measure itself and the inability to compare estimates of speech recognition across measures of a similar nature. Based on the implications described, it is of importance that follow on research aims to redress the drawbacks of the current methodology.

4.5 Study Limitations and Future Research Directions

Despite meticulous efforts to accurately implement the rigorous research methodology required when developing a new MST, several limitations arose in the current research that may challenge the utility of the results obtained. Each of these will be considered in the following sections with reference to how subsequent research may prevent such drawbacks.

4.5.1 The Use of Non-Optimised Stimulus Lists in the Babble Noise Conditions

A particularly important limitation that occurred in the current study involved the use of non-optimised stimulus lists in the babble noise conditions. As previously noted, a malfunction in the software, that was not detected until data collection had been completed, permitted the use of constant noise lists in the babble noise condition. This fault may have had a substantial effect on the data obtained for this condition and therefore requires urgent redress in subsequent research. In order to progress the development of the UCAMST it is essential that evaluation measurements are obtained for the babble noise condition using the stimulus lists designed for use with this type of noise. Based on the encouraging findings for the constant noise list stimuli, suggesting successful optimisation of the test material, one can expect equivalence between the lists and conditions of the UCAMST, following the removal of this error. Regardless, recall that evaluation of a new MST provides vital information regarding the equivalence of the test stimuli and the ability to compare results across conditions and other versions of the measure (Akeroyd et al., 2015). Thus, prior to investigating the use of the UCAMST in practice, it is recommended that this limitation be addressed by pairing the correct stimulus lists with the babble noise and retesting a sample of listeners with NH.

4.5.2 The Training Effect

As revealed by Hagerman (1984), MSTs are associated with a significant training effect, defined as the difference in SRT between the first and last lists performed (Wagener et al., 2003). Therefore, the number of trials completed, and thus familiarisation with the test material, has the ability to influence the individual's SRT. Accordingly, practice lists are administered when examining speech recognition using MSTs to enable performance to stabilise prior to the assessment procedure. As mentioned previously, the training effect has not yet been determined for the UCAMST, and was therefore estimated based on the accounts from previous literature (Akeroyd et al., 2015; Dietz et al., 2014; Hochmuth et al., 2012) for the practice phase of the current study. Although it is likely, based on the uniform methodology employed across the development of new MSTs, that this practice procedure would have been sufficient, it is possible that the training effect of the UCAMST differs to international MSTs. Should this be so, the validity of the current results could be questioned given that the participants would have been continuing to adjust to the task during the test procedure, thus influencing the estimates of slope and SRT obtained. It is pivotal that the training effect associated with the UCAMST is evaluated in subsequent research to ensure sufficient practice is provided prior to testing.

International MSTs have investigated this phenomenon through employing the adaptive procedure described by Brand and Kollmeier (2002). In this procedure two randomly interleaved tracks that converge at the 20% and 80% targets are utilised in obtaining an estimate of SRT for each list (Brand & Kollmeier, 2002). The number of lists selected for evaluation of the training period differed across international versions, however seven or eight lists of 20 sentences (i.e. double lists) were commonly used (Dietz et al., 2014; Hochmuth et al., 2012; Wagener et al., 2003). The literature documents that results from seven or eight adaptive measurements revealed the most important difference in performance to occur between the first and second lists, with an average improvement in SRT of 1.1 dB SNR (Dietz et al., 2014; Hochmuth et al., 2012). Thus, due to the non-significant improvements in SNR between the remaining trials, previous researchers have concluded that administering two lists of 20 sentences prior to testing is sufficient in order to obtain valid measurements of an individual's SRT (Dietz et al., 2014; Hochmuth et al., 2012; Wagener et al., 2003). It is advisable that future research follows the described procedure in order to determine the training effect for the UCAMST. Obtaining such information provides information vital to the progression of the UCAMST towards use in clinical and research settings.

4.5.3 **The Sample**

Sample Size. The current study aimed to employ a large sample of participants with NH to complete the protocol, however the number of participants able to be recruited was considerably lower than anticipated in each of the babble noise conditions. This limitation arose in two parts that require consideration. The first factor that lead to this smaller sample size was related to the exclusion of data. As discussed, a portion of the data was excluded from the final analyses due to the difficulty some participants experienced in completing the task. Examination of the data set revealed such data to be unrepresentative of the sample, consequently introducing bias into the data set, justifying its exclusion. In addition to this, a further error occurred in the software, involving the SNR at which the noise was presented. This error was not detected immediately, necessitating exclusion of the data from a further six participants. Time was the second factor that lead to the smaller sample size than required as stringent time constraints hindered the ability to initiate a further recruitment phase, following the removal of such data. Efforts to prevent the limitations that are intrinsically associated with the size of the sample, such as inadequate power to detect a genuine effect, would have been undertaken should this have been viable given the time permitted to complete this research. Therefore, it is advised that, when conducting similar evaluations in the absence of such time constraints, future researchers attempt to employ a greater number of participants in order to preserve the accuracy of the estimates of slope and SRT obtained.

Recruitment. A further limitation may have arisen in the current study due the procedure employed in the recruitment phase of the research. Advertisements were distributed throughout the University of Canterbury (Christchurch, NZ) community and, although the participant pool included several individuals from outside of this community, it is possible that this may have hampered the ability to recruit a larger sample. As mentioned above, a sample of 64 participants (i.e. 16 participants per condition) was recommended for the current research (Akeroyd et al., 2015), a number that, in retrospect, may have been more attainable with wider recruitment. However, as noted, time constraints made employing a further recruitment phase unviable for the current research. Accordingly, future work evaluating the babble noise stimulus lists should aim to implement recruitment procedures that will facilitate the research needs.

Generalisability. When examining the sample the issue of whether the results can be generalised beyond the sample also requires particular consideration. The aim of the evaluation process is to provide evidence of equivalence through a sample of listeners that are likely to represent the general performance expected for individuals with NH. While whether this was achieved by the current study cannot be ascertained at present, some features of the sample are of interest. First, on average, participants in each condition were between 20 and 30 years old. While no recommendations regarding the age of participants have been provided in the literature, the current sample captured performance from a relatively narrow demographic. This feature of the sample may, in part, be related to the recruitment procedure employed, limiting the ability to capture the performance of a wider age-range of listeners with NH. Another feature of the sample that may threaten the ability to generalise the findings to the wider NZ population is the gender balance. The participants involved in the current study were mostly female listeners, thus limiting the representation of performance in male listeners. It is interesting to note however that the underrepresentation of male listeners in this research is in line with findings suggesting poorer response rates to research advertisements, as compared to females (Patel, Doku, & Tennakoon, 2003). Further, a similar imbalance is observable in the evaluation procedures of previous MSTs (Ozimek et al., 2010; Wagener et al., 2003). Therefore, while this factor may not have had a direct impact on the current data, it is advisable that future researchers aim to employ a more representative sample in an attempt to preserve the generalisability of the findings.

4.6 Beyond the Current Study: Future Research Directions

When developing a new measure, the ultimate aim is for it to be incorporated into research and clinical test batteries. Accordingly, there are a number of areas, beyond the scope of the current research, that are of interest to the development of the UCAMST. A number of areas requiring attention in future research will be highlighted in the following sections.

4.6.1 **Cross-Validation with Other Speech Tests**

First, an area of research that follow on work should aim to address is the cross-validation of the UCAMST with other existing speech measures. Unlike the comparisons across MSTs conducted in the current study, the rationale behind this procedure is to gain insight into the information offered by various speech measures in an attempt to determine those that are complementary in practice. For the UCAMST, given the measures commonly incorporated into audiological test batteries in NZ clinics, cross-validation with the NZ CVC word lists (Purdy et al., 2000) is recommended. Another clinically available speech measure available in NZ is the QuickSIN (Killion et al., 2004). Despite not being routinely incorporated into the audiologic assessment in NZ, cross-validation of the QuickSIN (Killion et al., 2004) is recommended, due to the use of sentence stimuli in noise. Comparing the UCAMST with these two commercially available speech measures in future research would enable insight into the information that can be obtained from the UCAMST with respect to the NZ CVC word lists (Purdy et al., 2000) and the QuickSIN (Killion et al., 2004), and whether this is comparable between such measures. Based on the described work of Wilson et al. (2007a), examining the merit of various speech tests in combination, it is likely that such comparisons may provide valuable insight regarding the battery of speech tests that are most suitable for clinical use based on the information extracted and the amount of time required to obtain such information.

4.6.2 **Piloting with Individuals with HI**

As previously noted, the UCAMST has been administered to only participants with NH to date. Therefore, it is of interest to investigate the expected performance of listeners with HI on the UCAMST in future research. The variation in expected SRT among listeners with HI, as compared to listeners with NH, is well documented and, in part, has been ascribed to the influence of the spectrum of the masking noise (Peters et al., 1998). As previously described, when the background noise is different to the spectrum of the target stimulus, listeners with NH are able to take advantage of brief glimpses of the stimulus provided by temporal and spectral dips (Peters et al., 1998). Unfortunately, listeners with HI are thought to be unable to make use of these glimpses, thus effecting their SRT (Peters et al., 1998; Wilson et al., 2007b). Therefore, given the two noise types provided by the UCAMST software, investigation of the expected performance in each noise type for listeners with HI is warranted. The aim of such research would be to provide normative data, against which the performance of a given listener will be compared in order to determine the level of dysfunction faced in noise (Akeroyd et al., 2015).

4.6.3 **Examining the Application of the Auditory-Visual Mode**

A concurrent project (André, in progress) implemented the UCAMST in an investigation aimed at determining whether an individual's ability to make use of visual cues, in order to better understand speech, is related to HA outcomes. As described, the ability to combine sensory information is thought to be essential to efficient communication (Spehar et al., 2008; Tye-Murray et al., 2007a; Tye-Murray et al., 2014). The improvement in speech recognition that results from exploiting the information provided by both the auditory and visual modalities during communication has been termed “auditory-visual enhancement” (Tye-Murray et al., 2007b). Research has suggested that the ability to utilise this enhancement is largely preserved across the lifespan and is thought to be unaffected by HI (Spehar et al., 2008; Tye-Murray et al., 2007b). Accordingly, given that the UCAMST is the first MST to incorporate both auditory and visual presentation modes into its design, investigation into the application of such information was appropriate.

As noted, the most common audiologic intervention for HI is the provision of HAs, however it has been widely acknowledged in the literature that not all individuals that acquire HAs wear them routinely (Jerram & Purdy, 2001; Kelly-Campbell & Lessoway, 2015; Kochkin, 2000). HA disuse therefore typifies a major problem that rehabilitation audiologists need to strive to overcome during the prescription of HAs. A number of factors have been ascribed to positive HA outcomes, including self-perceived HI (Jerram & Purdy, 2001) and SNR loss (Allan, 2014; Robertson, Kelly-Campbell & Wark, 2012), however ambiguity around such factors remains. Regardless of the reasoning behind discontinued, or limited, HA use, investigation into tools that may possibly aid such negative outcomes was warranted.

Given the enhancement provided by the presence of both auditory and visual cues, alongside the prevalence of negative HA outcomes, André (in progress) aims to investigate the extent to which HA outcomes may be related to an individual’s auditory-visual integration skills. The study examines both new and experienced HA users’ performance in the auditory-alone and auditory-visual presentation modes of the UCAMST alongside self-reported HA outcomes (André, in progress). It is hoped that this study will provide information regarding the relationship between auditory-visual enhancement and HA outcomes and provide direction regarding the use of this tool clinically. It is possible that auditory-visual performance, as compared to auditory-alone performance, on the UCAMST may assist audiologists' in providing rehabilitation recommendations beyond HAs. Such recommendations may include perceptual training, in addition to informational counselling concerning communication strategies and/or lipreading (Tye-Murray, Sommers & Spehar, 2007c). Further, such information may aid the prevalence of negative HA outcomes through the ability to provide more realistic expectations to clients. Therefore, the purpose of André's (in progress) work is to contribute to the literature surrounding the clinical application of the UCAMST in audiologic rehabilitation.

4.6.4 **Investigating the Effect of Working Memory**

Working memory is thought to be particularly relevant in auditory speech processing due to the role it plays in both the storage and processing of the incoming message (Cervera et al., 2009). It has been reported that age-related working memory deficits, together with some degree of HI, may explain the increased deficits in communication experienced by elderly listeners (Rabbit, cited in Cervera et al., 2009). Research investigating this phenomenon, with regards to speech recognition, have indeed indicated that reduced memory capacity may influence estimates of SRT (Theunissen et al., 2009). In early research van Rooij and Plomp (1990) sought to examine the effect of memory on performance on speech recognition tasks including vowel and consonant identification, spondee word lists and sentence recognition in quiet and in noise. The results identified an association between reduced memory capacity and higher SRTs (van Rooij & Plomp, 1990). Based on this study the mediating effect of cognition on speech recognition could not be concluded, however these findings emphasise the need for consideration of an individual's memory capacity in conducting speech audiometry (van Rooij & Plomp, 1990). The authors concluded that in practice, speech recognition test procedures should be as brief as possible, particularly when working with elderly individuals, due to the increased likelihood of age-related cognitive changes being a factor in the results obtained (Cervera et al., 2009; van Rooij & Plomp, 1990). This suggestion is particularly important to speech audiometry, given that a large proportion of the population over 65 years of age are likely to be affected by HI, the need for employing accurate yet efficient measures of SRT into the audiological test battery becomes apparent (Newman & Sandridge, 2004). Further, due to the additional cognitive load associated with speech measures employing sentence stimuli, there is a need to consider the role of working memory in SRT estimation on such tasks (Cervera et al., 2009; McArdle et al., 2005; Wilson et al., 2007a). This is therefore applicable to the UCAMST, given that listeners are required to retain the sentence for enough time to be able to identify what was heard either verbally or by selecting the sentence in the closed set condition. Thus, future research quantifying the role of cognitive factors associated with speech recognition testing via MSTs may be helpful to the progression of the UCAMST. Such information may enable the use of the task in a clinical setting to be modified in order to accommodate such factors and may also provide a more complete representation of the deficits faced by clients in everyday communication. Accordingly, the ability to assist a client in establishing realistic rehabilitation goals may also be aided by such information.

The extent of this issue is not limited to speech recognition testing in older populations, however. The need for behavioural tests in paediatric audiology that meet certain criteria in order to be deemed valid in capturing speech recognition in children has also been acknowledged (Kosky & Boothroyd, 2003). Appropriately, upon development, the cognitive and attentional demands of tasks designed for use with children have required consideration (Kosky & Boothroyd, 2003). Due to the cognitive demands of the traditional MST (Hagerman, 1982) described, Ozimek, Kutzner, and Libiszewski (2012) modified the Polish MST (Ozimek et al., 2010) to enable this tool to be implemented in paediatric assessment procedures. The Polish Pediatric MST (PPMST) differs from the original Polish MST (Ozimek et al., 2010) in two key ways (Ozimek et al., 2012). First, since sentence tests for children generally employ short and simple stimuli, the number of columns in the base matrix was reduced from five to three in order to generate sentences of the fixed subject-verb-object structure (Ozimek et al., 2012). For example (Ozimek et 2012, p. 1123):

"babcia maluje dom"

(English translation: "Grandma is painting a house")

Secondly, the 16x3 base matrix required alteration in order to prevent the generation of nonsense sentences (Ozimek et al., 2012). To achieve this, the matrix was separated into four independent 4x3 sub-matrices (Ozimek et al., 2012). Through constructing the measure in this way its use became appropriate for children of seven years and over (Ozimek et al., 2012). Alongside consideration of the cognitive demands of paediatric speech recognition tools, the child's language and vocabulary competency also requires some thought (Kosky & Boothroyd, 2003). Accordingly, administration of the PPMST was adapted into a picture-point method to enable use with children aged three to six years old (Ozimek et al., 2012).

The importance of employing sentence intelligibility tasks for paediatric speech audiometry, instead of word recognition tasks, has been highlighted in the literature due to the greater amount of information regarding speech recognition that can be obtained from such measures (Bell & Wilson, 2001; Mendel, 2008). Accordingly, while the PPMST (Ozimek et al., 2012) addresses this need for the Polish language, the need for other language-specific versions remains. Given the merits of the MST format, such as its efficiency in estimating an individual's SRT, development of a paediatric version of the UCAMST would address this need in paediatric audiology in NZ. Therefore, investigating the ability to adapt the UCAMST into a version appropriate for use with paediatric populations may be of interest to future researchers.

4.7 Concluding Remarks

The current study contributed to a series of studies aimed at furthering the development of the UCAMST. The aim was to investigate the difficulty of the stimulus lists in order to determine equivalence. The results suggested that while the lists designed for use in the constant noise condition were of equal difficulty, the same was not true for the babble noise condition. Consequently, examination of the conditions in the UCAMST also revealed variations in the difficulty and sensitivity of each. From a clinical and research perspective these findings have important implications regarding the administration of the test lists, effecting the ability to employ the lists interchangeably. Additionally, it was of interest to the current research to determine whether the UCAMST is comparable to internationally published MSTs. Based on the findings throughout, it was unsurprising that differences between the measures were highlighted in these analyses, thus warranting further evaluation of the UCAMST stimulus lists. Follow on work should aim to address the drawbacks of the current research in order to permit concluding remarks regarding the equivalence, and therefore use, of the UCAMST to be proposed.

The MST has become popular in research surrounding speech audiometry in the last decade, due to the merits associated with this test format. It is thought that the MST is of benefit to both research and clinical settings, due to the efficiency and validity of SRT estimates, the ability to compare results across languages and clinics/research centres, and the large repertoire of stimulus sentences that make memorisation of the test materials unlikely. Since speech audiometry is an integral component of the audiological test battery, endeavouring to provide information beyond the audiogram that represents an individual's perceived deficits, the reason for such popularity becomes clear. Development of the UCAMST aims to provide NZ speech audiometry with this valuable tool. It is hoped that continuation of this study will be undertaken in order to progress the development of the UCAMST towards implementation in clinical and research practices in NZ.

REFERENCE LIST

** The figures cited throughout this work were reproduced in accordance with copyright laws, with permission granted by the copyright holders of published work **

Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., … & Kollmeier, B. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests: ICRA working group on multilingual speech tests. *International Journal of Audiology, 54* (Suppl. 2), 17-22. http://dx.doi.org/ 10.3109/14992027.2015.1030513

Allan, L. *Factors associated with hearing aid disuse in New Zealand/Aotearoa* (Master's Thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092/10775

André, E. (in progress). *Is auditory-visual integration ability a factor in hearing aid outcomes?* (Master's Thesis), University of Canterbury, Christchurch, New Zealand.

Arbogast, T. L. (2003). *The effect of spatial separation on informational and energetic masking of speech in normal-hearing and hearing-impaired listeners* (Doctoral Thesis), Boston University, USA. Retrieved from: http://search.proquest.com .ezproxy.canterbury.ac.nz/docview/305343579?pqorigsite=summon&accountid=1449

9

Arlinger, S. D. (1998). Clinical assessment of modern hearing aids. *Scandinavian Audiology, 27*(4), 50-53. http://dx.doi.org/10.1080/010503998420658

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and Hearing Research, 41*(3), 549-563. Retrieved from: http://search.proquest.com.ezproxy.canterbury.ac.nz/docview/232347131?pq-origsite=summon&accountid=14499

Beattie, R. C., Barr, T., & Roup, C. (1997). Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. *British Journal of Audiology, 31*, 153-164. http://dx.doi.org/10.3109/03005364000000018

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology, 12*(10), 514-522.

Bess, F. H., & Humes, L. E. (2008). *Audiology: The Fundamentals.* Philadelphia, USA: Lippincott Williams & Wilkins.

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology, 12*(3), 205-221. http://dx.doi.org/10.1080/09297040500488522

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27*, 32-48. http://dx.doi.org/10.1044/jshr.2701.32

Bird, P. A., & O'Beirne, G. A. (2015). Hearing loss in New Zealand – planning for the future. *New Zealand Medical Journal, 128*(1419), 6-8, from: http://www.nzma.org.nz /__data/assets/pdf_file/0003/43725/Ed-Bird-NZMJ1419.pdf

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology, 2*(1), 3-10. http://dx.doi.org/10.3109/00381796809075436

Boothroyd, A. (2008). The performance/intensity function: An underused resource. *Ear & Hearing, 29*(4), 479-491. http://dx.doi.org/10.1097/AUD.0b013e318174f067

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*(1). http://dx .doi.org/101-114. 10.1121/1.396976

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment, *Audiology, 34*(5), 260-284. http://dx.doi.org/10.3109 /00206099509071918

Brand, T. & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimation for psychophysics and speech intelligibility tests, *The Journal of the Acoustical Society of America, 111*(6). http://dx.doi.org/2801-2810.10.1121 /1.1479152

Carhart, R. T., & Young, L. (1976). Development of test procedures for evaluation of binaural hearing aids: A final report. *Bulletin of Prosthetics Research, 26*, 9-45. Retrieved from: http://www.rehab.research.va.gov/jour/76/13/2/carhart.pdf

Cervera, T. C., Soler, M. J. Dasi, C., & Ruiz, J. C. (2009). Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity, *Canadian Journal of Experimental Psychology, 63*(3), 216-226. http://dx.doi.org/10.1037 /a0014321

Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J. P., Abrams, H. B., Lesner, S., … Newman, C. W. (2007). A systematic review of health-related quality of life and hearing aids: Final report of the American Academy of Audiology task force on the health-related quality of life benefits of amplification in adults. *Journal of the American Academy of Audiology, 18*(2), 151-183.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *American, Speech-Language-Hearing Association, 23*(7), 493-500.

Cook, J. A., & Hawkins, D. B. (2006). Hearing loss and hearing aid treatment options. *Mayo Clinic Proceedings, 81*(2), 234-237. http://dx.doi.org/10.4065/81.2.234

Cox, R. M., Alexander, G. C., Gilmore, C. (1999). Development of the Connect Speech Test (CST). *Ear and Hearing, 8*(Suppl. 5), 119-126. http://dx.doi.org/10.1097/00003446-198710001-00010

Craik, F. I. M. (1998). Memory changes in normal ageing. *Current Directions in Psychological Science, 3*(5), 155-158. http://dx.doi.org/10.1111/1467-8721.ep10770653

Dietz, A., Buschermöhle, M., Aarnisalo, A. A., Vanhagen, A., Hyyrynen, T., Aaltonen, O., … & Kollmeier, B. (2014). The development and evaluation of the Finnish matrix sentence test for speech intelligibility assessment. *Acta Oto-Laryngologica, 134,* 728-737. http://dx.doi.org/10.3109/00016489.2014.898185

Dirks, D. D., Morgan, D. E., & Dubno, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*, 114-123. http://dx.doi.org/10.1044/jshd.4702.114

Donkelaar, H. J., & Kaga, K. (2011). The Auditory System. In H. J. ten Donkelaar (Ed). *Clinical Neuroanatomy: Brain Circuitry and Its Disorders.* New York: USA, Springer.

Francart, T., van Wieringen, A., & Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology, 50*, 2-13. http://dx.doi .org/10.3109/14992027.2010.505582

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet, 366*, 1111-1120. http://dx.doi.org/10.1016/S0140-6736(05)67423-5

Gelfand, S. A. (1998). Optimising the reliability of speech recognition scores. *Journal of Speech, Language, and Hearing Research, 41*(5), 1088-1102. Retrieved from: http:/ /search.proquest.com.ezproxy.canterbury.ac.nz/docview/232356121?accountid=1449 9

Gelfand, S. A. (2010). *Hearing: An Introduction to Psychological and Physiological Acoustics.* London, UK: Informa Healthcare.

Gilchrist, J. M., Jerwood, D., & Ismaiel, H. S. (2005). Comparing and unifying slope estimates across psychometric function models. *Perception and Psychophysics, 67*(7), 1289-1303. http://dx.doi.org/10.3758/BF03193560

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P., (2004). *New Zealand English: Its Origins and Evolution.* New York, USA: Cambridge University Press.

Grunditz, M., & Magnusson, L. (2013). Validation of a speech-in-noise test used for verification of hearing aid fitting. *Hearing Balance and Communication, 18,* 118-125. http://dx.doi.org/10.3109/21695717.2013.782135

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology, 11*(2), 79-87. http://dx.doi.org/10.3109/01050398209076203

Hagerman, B. (1984). Clinical measurement of speech recognition threshold in noise. *Scandinavian Audiology, 13*(1), 57-63. http://dx.doi.org/10.3109 /01050398409076258

Hall, S. J. (2006). *The Development of a New English Sentence in Noise Test and an English Number Recognition test* (Master's Thesis), University of Southampton, United Kingdom.

Hall, J. W. (2008). Diagnostic applications of speech audiometry. *Seminars in Hearing, 4*(3), 179-203. http://dx.doi.org/10.1055/s-0028-1091425

Hall, J. W. (2014). *Introduction to Audiology Today.* New Jersey, USA: Pearson.

Hällgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*(4), 227-237. http://dx.doi.org/10.1080/14992020500429583

Hamid, M. A., & Brookler, K. H. (2006). Speech audiometry. *Ear, Nose, & Throat Journal, 85*(12), 810-812.

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology, 51,* 536-544. http://dx.doi.org/10.3109 /14992027.2012.670731

Hope, R. V. (2010). *Towards the development of the New Zealand hearing in noise test (NZHINT)* (Master's thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from http://ir.canterbury.ac.nz/handle/10092/4815.

Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *Journal of the Acoustical Society of America, 125*(1), 442-446. http://dx.doi.org/10.1121/1.3037233

Hornsby, B. W. Y., Ricketts, T. A., & Johnson, E. E. (2006). The effects of speech and speechlike maskers on unaided and aided speech recognition in persons with hearing loss. *Journal of the American Academy of Audiology, 17*(6), 432-447. http://dx.doi .org/10.3766/jaaa.17.6.5

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., Verschuure, H., & Dreschler, W. A. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. *International Journal of Audiology, 53*(10), 760-763. http://dx.doi.org/10.3109/14992027.2014.920111

Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78*, 258-272.

Humes, L. E. (1999). Dimensions of hearing aid outcome. *Journal of the American Academy of Audiology, 10*(1), 26-39.

Hussain, S. S. M. (2008). *Conductive Hearing Loss* (Synopsis of Causation). Dundee, Scotland: Ministry of Defence.

Jamaluddin, S., & O'Beirne, G. A. (2015). *Development and evaluation of the auditory-visual matrix sentence test in Malay*: Paper presented at the Proceedings of the 7th Biennial Asia Pacific Conference of Speech, Language and Hearing for the Asia Pacific Society of Speech Language and Hearing, Guangzhou, China.

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormès., Frachet, B., Wouters, J., & van Wieringen, A. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *International Journal of Audiology, 51*(3), 164-173. http://dx.doi.org/10.3109 /14992027.2011.633568

Jerram, J. C. K., & Purdy, S. C. (2001). Technology, expectations, and adjustment to hearing loss: Predictors of hearing aid outcome. *Journal of the American Academy of Audiology, 12*(2), 64-79.

Kelly-Campbell, R. J., & Lessoway, K. (2015). Hearing aid and hearing assistance technology use in Aotearoa/New Zealand. *International Journal of Audiology, 54,* 308-315. http://dx.doi.org/10.3109/14992027.2014.979952

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 116*(4), 2395-2405. http://dx.doi.org/10.1121/1.1784440

King, S. M. (2010). *Development and Evaluation of a New Zealand Digit Triplet Test for Auditory Screening (Master's Thesis).* University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092/5679

Kochkin, S. (2000). MarkeTrak V: "Why my hearing aids are in the drawer": The consumers' perspective. *The Hearing Journal, 53*(2), 34-41. Retrieved from: http://users.clas.ufl .edu/mcolburn/Web-links/Nursing%20Lecture/ITD%20HA.PDF

Kollmeier, B., Warzybok, A., Hochmuth, H., Zokoll, M.A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology,* 1-13. http://dx.doi .org/10.3109/14992027.2015.1020971

Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility. *Journal of the Acoustical Society of America, 102*(4), 2412-2421. http://dx.doi.org/10.1121/1.419624

Kosky, C., & Boothroyd, A. (2003). Validation of an on-line implementation of the imitative test of speech pattern contrast perception (IMSPAC). *Journal of the American Academy of Audiology, 14*(2), 73-83.

Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology, 50*(5), 507-515. http://dx.doi.org/10.1111/j.1467-9450.2009.00747.x

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2), 467-477. http://dx.doi.org/10.1121/1.1912375

Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. *Frontiers in Psychology, 5*(639), 1-7. http://dx.doi.org/10.3389/fpsyg.2014.00639

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English “short” front vowel shift. *Language Variation and Change, 19*(1), 1-25. http://dx.doi.org/10.10170S0954394507070020

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing, 18*, 1-26. http://doi.dx.org/10.1177/2331216514537722

Mathers, C., Smith, A., & Concha, M. (2000). Global burden of hearing loss in the year 2000. *Global Burden of Disease, 18*, 1-30. Retrieved from: http://www.who.int/healthinfo /statistics/bod_hearingloss.pdf

Mattheyses,W., Latacz, L., & Verhelst, W. (2009). On the importance of audiovisual coherence for the perceived quality of synthesized visual speech. *EURASIP Journal on Audio, Speech, and Music Processing, 2009*(1), 1-12. http://dx.doi.org/10.1155 /2009/169819

McArdle, R.A., & Chisolm, T. H. (2009). Speech Audiometry. In J. Katz., L. Medwetsky., R. Burkard., & L. J. Hood (Eds). *Handbook of Clinical Audiology* (pp. 64-79). Philadelphia, USA: Lippincott Williams & Wilkins.

McArdle, R. A., Wilson, R. H., & Burks, C. A. (2005). Speech recognition in multitalker babble using digits, words and sentences. *Journal of the American Academy of Audiology, 16*(9), 453-470.

McClelland, A. (2014). *Refinement and normalisation of the University of Canterbury Auditory-Visual Matrix Sentence Test* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092 /10456.

Mendel, L. L. (2007). Objective and subjective hearing aid assessment outcomes. *American Journal of Audiology, 16*(2), 118-129. http://dx.doi.org/10.1044/1059-0889(2007/016)

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology, 47*(9), 546-553. http://dx.doi.org/10.1080/14992020802252261

Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., … & DeNino, L. A. (1990). Quality-of-life changes and hearing impairment: A randomized trial. *Annals of Internal Medicine, 113*(3)*,* 188-194.

Myerson, J., Spehar, B., Tye-Murray, N., Van Engen, K., Hale, S., & Sommers, M. S. (2016). Cross-modal informational masking of lipreading by babble. *Attention, Perception and Psychophysiology, 78*(1), 346-354. http://dx.doi.org/10.3758/s13414-015-0990-6

Niemeyer, W. (1976). Speech audiometry and fitting of hearing aids in noises. *International Journal of Audiology, 15*(5), 421-427. http://dx.doi.org/10.3109/00206097609071802

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of* *the Acoustical Society of America, 95*(2), 1085-1099. http://dx.doi.org/10.1121/1.408469

Newman, C. W., & Sandridge, S. A. (2004). Hearing loss is often undiscovered, but screening is easy. *Clevland Clinic Journal of Medicine, 71*(3), 225-232. http://dx.doi.org/10.3949/ccjm.71.3.225

O'Beirne, G. A., McGaffin, A. J., & Rickard, N. A. (2012). Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *International Journal of Pediatric Otorhinolaryngology, 76*(6), 777-782. http://dx.doi.org/10.1016/j.ijporl.2012.02.039

O'Beirne, G. A., Trouson, R. H., McClelland, A. D., Jamaluddin, S., Maclagan, M. A. (2015). Development of an auditory-visual matrix sentence test in New Zealand English. *12th European Federation of Audiological Societies Congress.* Istanbul, Turkey, 30 May 2015.

Olsen, W. O., Van Tasell, D. J., & Speaks, C. E. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear & Hearing, 18*(3), 175-188. http://dx.doi.org/10.1097/00003446-199706000-00001

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: A call to action. *Bulletin of the Word Health Organisation, 92*(5), 367-373. http://dx.doi.org/10.2471/BLT.13.128728.

Orchik, D. J., Krygier, K. M., & Cutts, B. P. (1979). A comparison of the NU-6 and W-22 speech discrimination tests for assessing sensorineural hearing loss. *Journal of Speech and Hearing Disorders, 44*(4), 522-527. http://dx.doi.org/10.1044/jshd.4404.522

Øygarden, J. (2009). *Norwegian Speech Audiometry* (Doctoral Thesis). Norwegian University of Science and Technology, Trondheim, Norway. Retrieved from: http://www.diva-portal.org/smash/get/diva2:214670/FULLTEXT01.pdf

Ozimek, E., Kutzner, D., & Libiszewski, P. (2012). Speech intelligibility tested by the Pediatric Matrix Sentence test in 3-6 year old children. *Speech Communication, 54*, 1121-1131. http://dx.doi.org/10.1016/j.specom.2012.06.001

Ozimek, E., Kutzner, D., Sęk, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology, 48*(7), 433-443. http://dx.doi.org/10.1080/14992020902725521

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49,* 444-454. http://dx.doi.org/10.3109/14992021003681030

Palm, W., & Glinos, I. A. (2010). Enabling patient mobility in the EU: Between free movement and coordination. In E. Mossialos., G. Permanand., R. Baeten., & T. K. Hervey (Eds). *Health systems and governance in Europe: The role of European Union and law policy* (pp. 509-560). Cambridge, England: Cambridge University Press.

Parving, A. (1991). The value of speech audiometry in hearing aid rehabilitation, *Scandinavian Audiology, 20*(3), 159-164. http://dx.doi.org/10.3109/01050399109074948

Patel, M. X., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment, 9*(3), 229-238. http://dx.doi.org/10.1192/apt.9.3.229

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Acoustical Society of America, 103*(1), 577-587. http://dx.doi.org/10.1121/1.421128

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America, 63*(2), 533-549. http://dx.doi.org.ezproxy.canterbury.ac.nz/10.1121/1.381753

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18*, 43-52. http://dx.doi.org/10.3109/00206097909072618

Puglisi, G. E., Astolfi, A., Prodi, N., Visentin, C., Warzybok, A., Hochmuth, S., & Kollmeier, B. (2014). *Construction and first evaluation of the Italian matrix sentence test for the assessment of speech intelligibility in noise.* Paper presented at the FORUM ACUSTICUM, Kraków: Poland. Retrieved from: http://www.fa2014.agh.edu.pl /fa2014_cd/article/RS/R03C_6.pdf

Purdy, S. C., Arlington, B., & Johnstone, C. (2000). Normative data for the New Zealand recording of the CVC (Revised AB) word lists, *New Zealand Audiological Society Bulletin, 10*(2), 20-29.

Schlauch, R. S., & Nelson, P. (2009). Puretone Evaluation. In J. Katz., L. Medwetsky., R. Burkard., & L. J. Hood (Eds). *Handbook of Clinical Audiology* (pp. 30-49). Philadelphia, USA: Lippincott Williams & Wilkins.

Schmiedt, R. A. (2010). The Physiology of Cochlear Presbycusis. In S. Gordon-Salant., R. D. Frisna, R. D., Popper, A. N., & Fay, R. R. (Eds). *The Ageing Auditory System* (pp. 9-34). New York, USA: Springer.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N* (L). *Journal of the Acoustical Society of America,* 118(5), 2775-2778. http://dx.doi.org/10.1121/1.2062650

Spehar, B. P., Tye-Murray, N., & Sommers, M. S. (2008). Intra-versus intermodal integration in young and older adults, *Journal of the Acoustical Society of America, 123*(5), 2858-2866. http://dx.doi.org/10.1121/1.2890748

Spencer, G. A. (2011). *Effects of speaker age on speech understanding and listener effort in older adults* (Master's Thesis). University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/bitstream/handle/10092/6343/Thesis _Fulltext.pdf?sequence=1&isAllowed=y

Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the Acoustical Society of America, 8*(2), 71-80.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212-215. http://dx.doi.org/10 .1121/1.1907309

Rebok, G. W., Smith, C. B., Pascualvaca, D. M., Mirsky, A. F., Anthony, B. J., & Kellam, S. G. (1997). Developmental changes in attentional performance in urban children from eight to thirteen years. *Child Neuropsychology, 3*(1), 28-46. http://dx.doi.org/10.1080 /09297049708401366

Robertson, M. A., Kelly-Campbell, R. J., & Wark, D. J. (2012). Use of audiometric variables to differentiate groups of adults based on hearing aid ownership and use. *Contemporary Issues in Communication Science and Disorders, 39*, 114-120.

Talbott, R. E., & Larson, V. D. (2008). Research needs in speech audiometry. *Seminars in Hearing, 4*(3), 299-308. http://dx.doi.org/10.1055/s-0028-1091432

Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal, 56*(1), 40-43. http://dx.doi.org/10.1097/01.HJ .0000293000.76300.ff

Theunissen, M., Swanepoel, D. W., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. *International Journal of Audiology, 48*(11), 743-757. http://dx.doi.org/10.3109/14992020903082088.

Trounson, R. H. (2012). *Development of the UC Auditory-Visual Matrix Sentence Test* (Master's Thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from: http://ir.canterbury.ac.nz/handle/10092/10348

Tye-Murray, N., Hale, S., Spehar, B., Myerson, J., & Sommers, M. S. (2014). Lipreading in school-age children: The roles of age, hearing status, and cognitive ability. *Journal of Speech, Language and Hearing Research, 57*(2), 556-565. http://dx.doi.org/10.1044 /2013_JSLHR-H-12-0273

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007b). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear & Hearing, 28*(5), 656-668. http://dx.doi.org/10.1097/AUD.0b013e31812f7185

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007a). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification, 11*(4), 233-241. http://dx.doi.org/10.1177/1084713807307409

Tye-Murray, N., Sommers, M. S., Spehar, B. (2007c). The effects of age and gender on lipreading abilities. *Journal of the American Academy of Audiology, 18*(10), 883-892. http://dx.doi.org/10.3766/jaaa.18.10.7

Tye-Murray, N., Sommers, M. S., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and younger adults in favourable and unfavourable conditions. *International Journal of Audiology, 47*(2), 31-37. http://dx.doi.org/10.1080/14992020802301662

Valente, M. (2009). *Pure-Tone Audiometry and Masking.* Oxfordshire, United Kingdon: Plural Publishing.

Van Engen, K., & Chandrasekaran, B. (2012). Sentence recognition as a function of the number of talkers in competing multi-talker babble. *Journal of the Acoustical Society of America, 132*(3), 1969. http://dx.doi.org/10.1121/1.4755263.

van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners II: Multivariate analyses. *Journal of the Acoustical Society of America, 88*(6). 2611-2642. http://dx.doi.org/ 10.1121/1.399981 van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111*(4), 1906-1916. http://dx.doi.org/10.1121/1.1456928

Versfeld. N. J., Daalder, L., Festen. J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of the Acoustical Society of America, 107*(3), 1671-1684. http://dx.doi.org/10.1121/1.428451

Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with NH and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology, 44,* 144-156. http://dx.doi.org/10.1080/14992020500057517

Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Old- enburger Satztests (Development and evaluation of a German speech intelligibility test. Part III: Evaluation of the Oldenburg sentence test). *Z Audiology, 38,* 86–95. Retrieved from: http://www.z-audiol.de/filebase/oeffentlich/1999/1999_03_Original_086-095.pdf

Wagener, K., Hochmuth, S., Ahrlich, M., Zokoll, M. A., Laan, V. D., & Kollmeier, B. (2014). Der weibliche Oldenburger Satztest (The female oldenburg sentence test). *Proceedings of 17th Congress of the German Society of Audiology, Germany,* *Oldenburg*. Retrieved from: https://translate.google.co.nz/translate?hl=en&sl=de&u=http://www.uzh.ch/orl/dga2014/programm/wissprog/Wagener.pdf&prev=search

Wagener, K., Josvassen, J. L., & Ardenkjoer, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42*(1), 10-17. http://dx.doi.org/10.3109/14992020309056080

Walden, B. E., Demorest, M. E., & Helper, E. L. (1984). Self-report approach to assessing benefit derived from amplification. *Journal of Speech and Hearing Research, 27*(1), 49-56. http://dx.doi.org/10.1044/jshr.2701.49

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test. *International Journal of Audiology, 54*(suppl. 2), 35-43. http://dx.doi.org/10.3109/14992027.2015.1020969

Welberg, L. (2008). Auditory system: Making hairs for hearing. *Nature Reviews Neuroscience, 9*(10), 1-2. http://dx.doi.org/10.1038/nrn2496

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14*(9), 453-470.

Wilson, R. H., Carnell, C. S., & Cleghorn, A. L. (2007b). The words-in noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology, 18*(6), 522-529. http://dx.doi.org.ezproxy.canterbury.ac.nz /10.3766/jaaa.18.6.7

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007a). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with NH and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 50*(4), 844-856. http://dx.doi.org/10.1044/1092-4388(2007/059)

Wilson, R. H., Trivette, C. P., Williams, D. A., & Watts, K. L. (2012). The effects of energetic and informational masking on the Words-in-Noise Test (WIN). *Journal of the American Academy of Audiology, 23*(7), 522-533. http://dx.doi.org/10.3766 /jaaa.23.7.4

World Health Organisation, (2001). *International Classification of Functioning, Disability, and Health, ICF.* Geneva: World Health Organization.

Wong, L. L. N., & Soli, S. D. (2005). Development of the Cantonese hearing in noise test (CHINT). *Ear and Hearing, 26*(3), 276-289. http://dx.doi.org/10.1097/00003446-200506000-00004

Zeng, F. G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research, 49*(2), 367-380. http://dx .doi.org/10.1044/1092-4388(2006/029)

Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-noise tests for multilingual hearing screening and diagnostics. *American Journal of Audiology, 22*(1), 175-178. http://dx.doi.org /10.1044/1059-0889(2013/12-0061).

Appendices

APPENDIX A: *ETHICAL APPROVAL*

Letter of ethical approval, University of Canterbury Human Ethics Committee.

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2014/49

11 May 2015

Jessica Stone
Department of Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Jessica

Thank you for your request for an amendment to your research proposal "Naturalisation and normalisation of the UC auditory-visual matrix sentence test" as outlined in your email dated 4 May 2015.

I am pleased to advise that this request has been considered and approved by the Human Ethics Committee.

Yours sincerely

Lindsey MacDonald
Chair, Human Ethics Committee

APPENDIX B: *RECRUITMENT*

B.1 Study advertisement utilised during recruitment.

Volunteers Needed!

To help develop a NZ Auditory-Visual Speech Test

We are looking for participants that:

- Are 18 years of age or older
- Have **normal hearing**
- Are **native speakers of New Zealand English**
- Have **no chronic dexterity issues**

We are developing an exciting new speech test called the UC Auditory-Visual Matrix Sentence Test (UCAMST). The UCAMST will use both auditory and visual stimuli to diagnose hearing loss in NZ.

This study will take place at the University of Canterbury Speech & Hearing Clinic at Creyke Road, Ilam, throughout 2015.

You will be needed for one session of two hours, and during this time you will:

- **Get a free hearing check**
- **Receive a $10 petrol voucher**
- **Get a first-hand look at this new speech test**
- **Help to develop an exciting new speech test for use in NZ clinics!**

For more information, or to be involved in this project, please contact **Jessica Stone** at **jessica.stone@pg.canterbury.ac.nz** or text/call **027 6440031**

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

Speech test
Jessica – 027 6440031
jessica.stone@pg.canterbury.ac.nz

B.2 Email invitation utilised during recruitment.

Hi everyone,

VOLUNTEERS ARE NEEDED!

The UC Auditory-Visual Matrix Sentence Test is an exciting new speech test that uses both auditory and visual cues in the diagnosis of hearing loss. Our goal is to further develop the test for use in NZ audiology clinics in the future.

If you:

-Are 18 years of age or older

-Have normal hearing

-Are a native speaker of NZ English

-Have no chronic dexterity issues

Then I would like to hear from you!

This study will take place at the University of Canterbury Speech & Hearing Clinic at Creyke Road, Ilam, throughout 2015.

You would be needed for one session of one hour and during this time you will:

-Get **a free hearing check**

-Receive **a $10 petrol voucher**

-See this exciting new speech test first hand and **help to develop it** further so it may be used in NZ audiology clinics in the future

For more information please email Jessica Stone at **jessica.stone@pg.canterbury.ac.nz** or phone/txt **027 6440031**

Thank you for reading!

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee

APPENDIX C: *INFORMED CONSENT*

C.1 Information sheet given to participants in the current study (page 1 of 2).

Information Sheet

Full Project Title:	**Evaluation of the University of Canterbury Auditory-Visual Matrix Sentence Test**
Principal Researcher:	Jessica Stone, MAud Student (2nd year) Department of Communication Disorders
Research Supervisor:	Associate Professor Greg O'Beirne Department of Communication Disorders
Associate Supervisor:	Dr. Rebecca Kelly-Campbell, Senior Lecturer Department of Communication Disorders

This study is part of a project to develop an auditory-visual speech test in NZ English to supplement the information gathered from other tests typically used in audiology. The study aims to assess the difficulty of the sentence lists to ensure each of the lists are of equal difficulty.

The test will take place at the University of Canterbury Speech and Hearing Clinic

To be eligible to participate, you must:
-be 18 years of age or older
-have normal hearing
-be a native NZ English speaker
-have no current middle ear pathology (i.e. ear infections)

Prior to any testing, you will be asked for a history of your hearing health, which ethnic group you belong to, and your ears will be examined. You will then undergo a hearing check to determine your hearing ability (i.e. whether you have normal hearing or whether you have a hearing impairment and, if so, to what degree), alternatively if you have an audiologist-completed audiogram dated within six months you will not be required to undergo this check. I will inform you of the results of your hearing test and, if you would like me to, I can write a letter summarising the results if you would like to follow up on this with your GP or an audiologist. In the event of an unexpected diagnosis of a hearing loss, a full audiological assessment will be offered at the University of Canterbury Speech and Hearing Clinic free of charge. If you choose to follow up with your GP, this will be at your own expense. If a conductive hearing loss were to be identified during the hearing check, you will receive a $10 fuel voucher for your time.

Following these checks, the study will begin. Short sentences being read in noise will be presented to you. The words will change in loudness and may at times be difficult for you to hear. After each sentence has been read, you will be asked to identify what you thought you heard. The study will require a maximum of 2 hours of your time.

This study is being carried out as part of a Masters of Audiology. The information I obtain from you will be used in further development of this test so that it may be used as a diagnostic tool.

C.1 Information sheet given to participants in the study (page 2 of 2).

I am happy to answer any queries you may have. My telephone and email details are provided in case you have any questions at a later date. In recognition of the time and effort involved on your behalf, you will receive an honorarium of $10, as well as a free hearing check.

I have provided a consent form for you to sign prior to participating in this study.

Signing this indicates your understanding that the data collected in this study will not be anonymous, but it will be confidential, and only viewed by people directly involved in this study (those listed at the top of the first page). Participation is voluntary and you have the right to withdraw at any stage without penalty. If you withdraw, I will remove all of the information relating to you.

The project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

For your own reference, please take this form away with you.

With thanks,

Jessica Stone
2nd year MAud Student
Department of Communication Disorders
University of Canterbury
Email: jessica.stone@pg.canterbury.ac.nz
Phone: 0276440031

Greg O'Beirne, PhD
Primary research supervisor & Associate Professor in Audiology
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: gregory.obeirne@canterbury.ac.nz
Phone: +64 3 364 2987 ext. 7085

Rebecca Kelly-Campbell
Secondary research supervisor & Senior Lecturer in Audiology
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: rebecca.kelly@canterbury.ac.nz
Phone: +64 3 364 2987 ext. 8327

Alternatively, if you have any complaints, please contact the Chair of the University of Canterbury Human ethics committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz), phone: +64 3 364 2987.

C.2 Consent form signed by all participants involved in the study.

Consent Form for Persons Participating in Research Studies

Full Project Title: *Evaluation of the University of Canterbury Auditory-Visual Matrix Sentence Test*

I have read and understand the Information Sheet.

I, ______________________ agree to participate in this project according to the conditions in the Information Sheet. I will be given a copy of Information Sheet and Consent Form to keep.

The researcher has agreed not to reveal the participant's identity and personal details if information about this project is published or presented in any public form.

I agree that research data gathered in this study may be published and used in future studies. I provide consent for this publication and the re-use of the data with the understanding that my name or other identifying information will not be used.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.

I understand the risks associated with taking part and how they will be managed.

I understand that I can contact the researcher or supervisor for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

I would like to receive a report on the findings of the study at the conclusion of the study (please tick one):

Yes ☐ **No** ☐

If yes, please provide a contact email and/or postal address below:

....

By signing below, I agree to participate in this research project.

Signature **Date**

....

Note: All parties signing the Consent Form must date their own signature. **Please return the consent form to the researcher before you actively participate in this research.**